

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

**Інститут прикладного системного аналізу Кафедра математичних
методів системного аналізу**

«На правах рукопису»

УДК 004.942; 519.254

«До захисту допущено»

Завідувач кафедри

_____ О.Л. Тимошук

«___» _____ 20__р.

Магістерська дисертація

на здобуття ступеня магістра

зі спеціальності 124 Системний аналіз

**на тему: «Математичне моделювання і прогнозування нелінійних
нестационарних процесів в економіці та фінансах»**

Виконала:

Студентка II курсу, групи КА-92мп

Ревуцька Людмила Олександрівна

Керівник:

професор кафедри ММСА,

д.т.н., проф.,

Бідюк П.І.

Рецензент:

декан ФІОТ КПІ ім. Ігоря Сікорського,

д.т.н., проф.,

Теленик С.Ф.

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць інших
авторів без відповідних посилань.

Студентка _____

Київ
2020

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
«Інститут прикладного системного аналізу»
Кафедра математичних методів системного аналізу

Рівень вищої освіти – другий (магістерський)

Спеціальність – 124 «Системний аналіз»

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ О. Л. Тимошук
«__» _____ 20__ р.

ЗАВДАННЯ
на магістерську дисертацію студентці

Ревуцькій Людмилі Олександрівні

1. Тема дисертації: «Математичне моделювання і прогнозування нелінійних нестационарних процесів в економіці та фінансах»,

науковий керівник дисертації Бідюк Петро Іванович, д.т.н., професор,

затверджені наказом по університету від «02» листопада 2020 р. № 3182-с.

2. Термін подання студенткою дисертації 14 грудня 2020 року.

3. Об'єкт дослідження: нелінійні нестационарні процеси в економіці та фінансах, представлені статистичними даними стосовно їх розвитку.

4. Предмет дослідження: математичні моделі аналізу даних з метою моделювання і прогнозування розвитку нелінійних нестационарних процесів; множини критеріїв для аналізу адекватності моделей та оцінювання якості прогнозів.

5. Перелік завдань, які потрібно розробити:

- 1) виконати огляд існуючих типів нелінійних нестационарних процесів (ННП);
- 2) виконати аналіз сучасних методів, які використовуються для моделювання і прогнозування ННП в економіці та фінансах;
- 3) виконати огляд деяких результатів відомих досліджень;

- 4) вибрати методи для моделювання і прогнозування ННП в економіці та фінансах;
- 5) вибрати методика для моделювання ННП;
- 6) розробити програмний продукт для моделювання і прогнозування ННП; виконати обчислювальні експерименти з метою моделювання;
- 7) виконати порівняльний аналіз отриманих результатів і зробити висновки;
- 8) виробити рекомендації щодо моделювання ННП.

6. Орієнтовний перелік ілюстративного матеріалу:

- 1) огляд процесів часових рядів;
- 2) класифікація моделей для прогнозування нелінійних нестационарних процесів;
- 3) ряди часових рядів обраних процесів;
- 4) результати прогнозування часових рядів;
- 5) таблиці у розділі стартап-проекту.

7. Орієнтовний перелік публікацій

- 1) Ревуцька Л.О., Бідюк П.І. Математичне моделювання і прогнозування нелінійних нестационарних процесів в економіці та фінансах. Проблеми інформатизації: матеріали VIII міжнародної наук.-техн. конф. (Черкаси, 26-27 листопада 2020 р.). Черкаси: ФОП Петров В.В., 2020. С. 21
- 2) Ревуцька Л.О., Бідюк П.І. Моделі нелінійних нестационарних процесів у фінансах та економіці. Університетський науковий збірник «Системні науки та кібернетика» – К.: NTUU «КРІ», 2020. С.159-173 (прийнято до друку).

8. Дата видачі завдання 01 вересня 2020 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Пошук літератури та поверхневе дослідження за тематикою «Математичне моделювання та прогнозування динамічних систем».	01.09.2020 – 06.09.2020	
2.	Огляд основних типів процесів та їхні особливості. Огляд відомих досліджень в області	07.09.2020 – 14.09.2020	

	прогнозування нелінійних нестационарних процесів.		
3.	Огляд існуючих підходів до моделювання і прогнозування нелінійних нестационарних процесів.	14.09.2020 – 28.09.2020	
4.	Вибір та опис методики моделювання нелінійних нестационарних процесів (методика Бокса-Дженкінса). Опис критеріїв для аналізу якості даних, адекватності моделей та якості оцінок прогнозів. Огляд тестів для аналізу даних на лінійність, інтегрованість та гетероскедастичність.	28.09.2020 – 12.10.2020	
5.	Збір та попередня обробка статистичних даних для виконання обчислювальних експериментів.	12.10.2020 – 26.10.2020	
6.	Вибір методів для виконання досліджень, підготовка до моделювання.	26.10.2020 – 2.11.2020	
7.	Виконання обчислювальних експериментів. Аналіз адекватності отриманих результатів.	02.11.2020 – 23.11.2020	
8.	Розробка плану стартап-проекту.	23.11.2020 – 30.11.2020	

Студентка _____

Л.О. Ревуцька

Науковий керівник дисертації _____

П.І. Бідюк

РЕФЕРАТ

Магістерська дисертація: 103 с., 20 рис., 28 табл., 57 джерел, 1 додаток.

Актуальність теми. Мінімізація фінансових ризиків та необхідність прийняття обґрунтованих управлінських рішень вимагають вдосконалення існуючих та пошуку нових підходів для реалізації високоякісних прогнозів.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась згідно з планом науково-дослідних робіт кафедри математичних методів системного аналізу Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Мета дослідження. Підвищення адекватності математичних моделей нелінійних нестационарних процесів в економіці та фінансах і якості оцінок їх прогнозів за рахунок створення нових математичних моделей.

Об'єкт дослідження. Нелінійні нестационарні процеси в економіці та фінансах, представлені статистичними даними стосовно їх розвитку.

Предмет дослідження. Математичні моделі аналізу даних з метою моделювання і прогнозування розвитку нелінійних нестационарних процесів; множини критеріїв для аналізу адекватності моделей та оцінювання якості прогнозів.

Методи дослідження. Використання регресійного підходу та методу групового урахування аргументів, метод найменших квадратів для оцінки параметрів моделі.

Наукова новизна отриманих результатів. Розроблено програмний продукт для моделювання та прогнозування нелінійний нестационарних процесів.

Апробація результатів дисертації. За матеріалами дисертаційної роботи опубліковано 1 тези конференції та подано до друку 1 статтю.

НЕЛІНІЙНІСТЬ, НЕСТАЦІОНАРНІСТЬ, ЧАСОВИЙ РЯД, КОРЕЛЯЦІЯ, ПРОГНОЗ, МОДЕЛЬ, РЕГРЕСІЙНИЙ ПІДХІД

ABSTRACT

The theme: Mathematical modeling and forecasting of nonlinear nonstationary processes in economics and finance.

Master's thesis: 103 p., 20 fig., 28 tab., 57 sources, 1 appendix.

Topic Relevance. Minimizing financial risks and the need to make sound management decisions require improving existing and finding new approaches to implementing high-quality forecasts.

Thesis connection to scientific programs, plans, and topics. The thesis was prepared according to the scientific research plan of the Department of Mathematical Methods of System Analysis of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute.

Research goal. Improving the adequacy of mathematical models of nonlinear nonstationary processes in economics and finance and the quality of estimates of their forecasts by creating new mathematical models.

Object of research. Nonlinear non-stationary processes in economics and finance, represented by statistics on their development.

Subject of research. Mathematical models of data analysis for the purpose of modeling and forecasting the development of nonlinear nonstationary processes; sets of criteria for analyzing the adequacy of models and assessing the quality of forecasts.

Methods of research. Using a regression approach and the method of group consideration of arguments, the method of least squares to estimate the parameters of the model.

Scientific contribution. A software product for modeling and forecasting of nonlinear nonstationary processes has been developed.

NONLINEARITY, NONSTATIONARY, TIME SERIES, CORRELATION, FORECAST, MODEL, REGRESSION APPROACH

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ І ТЕРМІНІВ	9
ВСТУП	10
РОЗДІЛ 1 АКТУАЛЬНІСТЬ ПРОБЛЕМИ ДОСЛІДЖЕНЬ, ДЕЯКІ ІСНУЮЧІ РЕЗУЛЬТАТИ.....	11
1.1 Аналіз актуальності проблеми моделювання і прогнозування нелінійних нестационарних процесів в економіці та фінансах	11
1.2 Теоретичні відомості про процеси	12
1.3 Існуючі підходи до моделювання і прогнозування нелінійних нестационарних процесів та огляд результатів досліджень	16
1.3.1 Параметричні моделі.....	17
1.3.2 Непараметричні моделі.....	21
1.4 Деякі сучасні програми для моделювання і прогнозування ННП	24
1.4.1 Statistica	25
1.4.2 EViews	27
1.4.3 SAS.....	28
1.4.4 Python (бібліотека SciPy)	29
Висновки до розділу 1 та постановка задачі дослідження	30
РОЗДІЛ 2 ВИБРАНІ МЕТОДИ І ПІДХОДИ ДО МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ФІНАНСАХ.....	32
2.1 Регресійний підхід	32
2.2 Метод групового урахування аргументів (МГУА).....	33
2.3 Методика моделювання нелінійних нестационарних процесів	35
2.3.1 Методика побудови математичних моделей часових рядів	35
2.3.2 Множини критеріїв для аналізу якості даних, адекватності моделей та якості оцінок прогнозів	42
2.3.3 Тести для аналізу даних на інтегрованість та гетероскедастичність ...	47
Висновки до розділу 2	50
РОЗДІЛ 3. ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ	51
3.1. Опис програмного продукту	51

3.2. Вибір даних та моделювання і прогнозування	51
Висновки до розділу 3	69
РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ	70
4.1 Опис ідеї проекту	70
4.2 Технологічний аудит ідеї стартап-проекту	72
4.3 Аналіз ринкових можливостей запуску стартап-проекту.....	73
4.4 Аналіз ринкової стратегії проекту.....	81
4.5 Розроблення маркетингової програми стартап-проекту	85
Висновки до розділу 4	91
ВИСНОВКИ.....	92
ПЕРЕЛІК ПОСИЛАНЬ	93
ДОДАТОК А ЛІСТИНГ ПРОГРАМИ	98

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, СКОРОЧЕНЬ І ТЕРМІНІВ

АР – авторегресія

АРКС – авторегресія з ковзним середнім

ММП – метод максимальної правдоподібності

МНК – метод найменших квадратів

МАП – максимальна абсолютна похибка

МіАП – мінімальна абсолютна похибка

МП – максимальна правдоподібність

НКФ – нелінійна кореляційна функція

ННП – нелінійний нестационарний процес

ПП – програмний продукт

САПП – середня абсолютна похибка в процентах

СКП – середньоквадратична похибка

СП – середня похибка прогнозу

СПП – середня похибка в процентах

CPI – Consumer price index (індекс споживчих цін)

R^2 — коефіцієнт детермінації

ВСТУП

Для передбачення майбутніх станів та процесів у соціальних, екологічних, технічних, фінансових, економічних та інших сферах діяльності здійснюється прогнозування на коротко- або середньостроковий термін.

Щоб прогнозувати і відслідковувати стан системи необхідні дані часових рядів, оскільки вони показують динамічну поведінку і причинно-наслідкові зв'язки, які є основою процесів. Для більшості сучасних систем, особливо в економіці та фінансах, притаманна нелінійна динаміка процесів з різними тенденціями, тобто ці процеси демонструють нестационарність. Нестационарність означає, що певні параметри (математичне сподівання, дисперсія або коваріація) досліджуваного процесу змінюються у часі на інтервалах дослідження.

Нелінійні нестационарні процеси також зустрічаються в перехідній економіці та у фінансовій діяльності. До таких процесів можна віднести різні макроекономічні показники, наприклад, інфляцію, ВВП, курси валют, ціни активів. Цей процес характеризується високою динамікою, а також позитивним і негативним трендом на різних етапах розвитку.

Нелінійний підхід у моделюванні є найбільш близьким до дійсності, оскільки при цьому способі враховуються всі можливі зміни параметрів. Проте в той же час побудова нелінійних нестационарних моделей і прогнозування процесів може бути досить складним завданням. Це створює серйозну проблему для точності прогнозування, особливо якщо такі процеси зазнають швидких і часом погано передбачуваних зовнішніх змін. Тим не менш, у літературі увага здебільшого зосереджена на прогнозуванні лінійних та стаціонарних процесів, що є зручним і простішим варіантом виконання досліджень [1].

Вдосконалення обчислювальної техніки і комп'ютерне моделювання в більшій мірі компенсує зростаючі труднощі, це сприяє розвитку різних підходів в області виконання аналізу економічних та фінансових процесів.

РОЗДІЛ 1 АКТУАЛЬНІСТЬ ПРОБЛЕМИ ДОСЛІДЖЕНЬ, ДЕЯКІ ІСНУЮЧІ РЕЗУЛЬТАТИ

1.1 Аналіз актуальності проблеми моделювання і прогнозування нелінійних нестационарних процесів в економіці та фінансах

Нелінійна економіка бере свій початок в 1940-х роках, коли угорський економіст Ніколас Калдор встановив, що модель з лінійними функціями для прогнозування інвестування та заощаджень не може коректно пояснити торговий цикл. При моделюванні заощаджень та інвестицій як лінійної функції зайнятості, якщо функція заощаджень була б більш стрімкою, тоді модель демонструвала б вищу стабільність, ніж це може бути насправді в реальному світі. З іншого боку, якщо інвестиційна функція була би більш стрімкою, то система завжди би прямувала до стану гіперінфляції з повною зайнятістю або до стану повного колапсу з нульовою зайнятістю [2]. Оскільки жоден результат не може бути виправданим, Калдор припустив, що інвестиційна функція $I(x)$ і функція заощаджень $S(x)$ не можуть бути лінійними, принаймні на усьому діапазоні можливих значень [2].

Застосування нелінійних методів в економіці після цього зауваження було дуже обмеженим, і Калдор пізніше стверджував, що це пояснюється тим, що раніше сприймалось як належне те, що економіка завжди наближається до стану рівноваги або близька до нього [3]. Тому економісти, як правило, покладалися на порівняльні статистичні методи, навіть коли відповідна лінійна модель була нестабільною за прийнятних значень параметрів – як у випадку з лінійною моделлю попиту та пропозиції, яка нестабільна за реальних умов, коли ціна пропозиції перевищує попит.

Хоча Калдор вказував на необхідність застосування нелінійного аналізу в економіці, піонером нелінійних моделей в економіці був Річард Гудвін. Гудвін розробив безліч нелінійних моделей та реалізував модель циклу класової боротьби, представивши її як систему хижака-жертви.

Зараз існує безліч нелінійних динамічних моделей макроекономіки, мікроекономіки та явищ фінансового ринку, в тому числі сукупний діловий цикл, окремий ринковий цикл та крахи фондового ринку. Також є множина економетричних методів для перевірки економічних та фінансових часових рядів на нелінійні структури даних (наприклад, показник Херста, статистика Брока-Дешерта-Шейнкмана, авторегресія плавного переходу). Проте їхня надійність, при визначенні наявності нелінійних причинно-наслідкових структур лінійних часових рядів з шумами, обмежена.

З точки зору статистики, часові ряди можна вважати такими, що описують нелінійні динамічні системи, якщо вони відзначаються такими характеристиками, як: аперіодичність, асиметричні цикли, мультимодальність, нелінійні причинно-наслідкові зв'язки між змінними та ін. [4].

Ефективне прогнозування майбутніх станів складної системи за часовими рядами залишається одним із складних та актуальних завдань сучасної науки, головним чином через різноманітні комбінації нелінійної та нестационарної динамічної поведінки, що демонструють ці системи [5].

Тому поглиблене дослідження, математичне моделювання та прогнозування нелінійних нестационарних процесів є актуальним завданням в сучасній економіці та фінансах зважаючи на зв'язок з такими фундаментальними проблемами в Україні та світі, які складно спрогнозувати. Особливо актуальним є управління фінансовими процесами на українського ринку, де фінансовий ризик-менеджмент перебуває на початковій стадії свого розвитку.

1.2 Теоретичні відомості про процеси

Всі досліджувані процеси умовно можна поділити на лінійні та нелінійні (тобто динамічні, але з лінійною та нелінійною динамікою) (Рис. 1.1). Процес вважається лінійним, якщо він описується лінійною моделлю з достатнім

ступенем адекватності, в протилежному випадку – процес нелінійний. Зазвичай лінійні процеси є стійкими, відносно легко прогнозованими і за допомогою минулих даних є можливість визначити тенденцію, що зручно при прогнозуванні. Лінійні процеси поділяються на стаціонарні та нестаціонарні (Рис. 1.2). При цьому до лінійних нестаціонарних відносяться процеси з лінійними трендами.

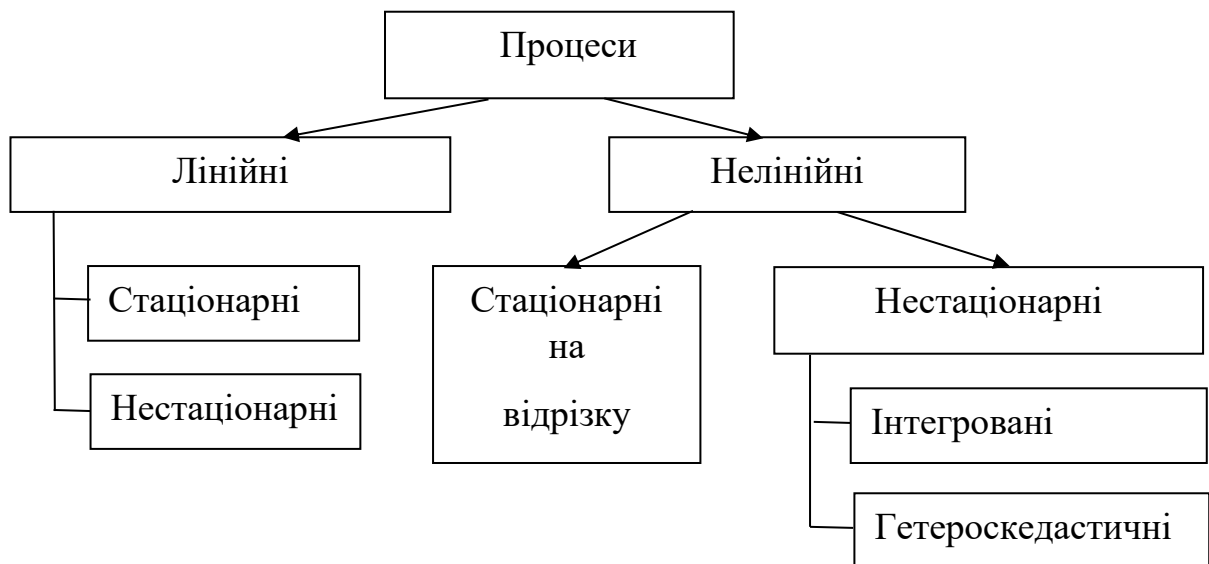


Рисунок 1.1 – Класифікація лінійних і нелінійних процесів

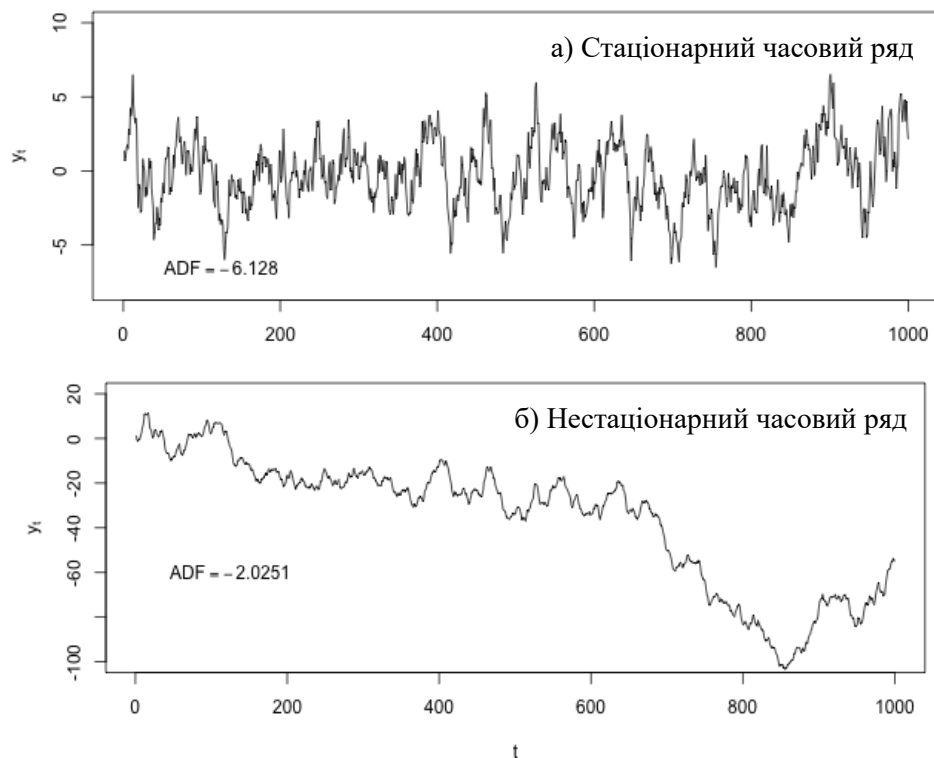


Рисунок 1.2 – Графіки процесів часових рядів [6]

Також в процесах розрізняють сильну і слабку стаціонарність. Слабку стаціонарність ще називають: стаціонарністю по коваріації або стаціонарністю другого порядку. Стохастичний процес із скінченним середнім та дисперсією є стаціонарним по коваріації якщо для всіх k і $k-s$ ($k = 0, 1, 2, \dots$; $s = 0, 1, 2, \dots$) виконуються такі умови:

- 1) математичне сподівання не залежить від часу:

$$E[y(k)] = E[y(k-s)] = \mu = \text{const}; \quad (1.1)$$

- 2) дисперсія залишається постійною для всього часового інтервалу, на якому розглядається процес:

$$E\{[y(k) - \mu]^2\} = E\{[y(k-s) - \mu]^2\} = \sigma^2 = \text{const}, \quad (1.2)$$

або

$$\text{var}[y(k)] = \text{var}[y(k-s)] = \sigma^2 = \text{const}; \quad (1.3)$$

- 3) автоковаріація залишається незмінною в часі для всього часового інтервалу:

$$\begin{aligned} E\{[y(k) - \mu][y(k-s) - \mu]\} = \\ = E\{[y(k-j) - \mu][y(k-j-s) - \mu]\} = \gamma(s) = \text{const}, \end{aligned} \quad (1.4)$$

або

$$\text{cov}[y(k), y(k-s)] = \text{cov}[y(k-j), y(k-j-s)] = \gamma(s) = \text{const}. \quad (1.5)$$

Сильна стаціонарність не вимагає, щоб математичне сподівання та дисперсія були постійними величинами. Таким чином, слабка стаціонарність накладає більш строгі обмеження на поведінку випадкового процесу, ніж сильна

стаціонарність. Отже, якщо хоча б один статистичний параметр (математичне сподівання, дисперсія або коваріація) процесу змінюється у часі на інтервалі дослідження, тобто не виконується хоча б одна з умов (1.1 – 1.3), то процес вважається нестаціонарним [7].

Нелінійні нестаціонарні процеси можна розділити на такі підгрупи: інтегровані процеси і гетероскедастичні.

Інтегрований процес – нестаціонарний процес, математичне сподівання якого змінюється у часі на інтервалі дослідження, тобто процес із трендом. Під трендом будемо розуміти загальний довгостроковий напрям розвитку процесу, порядок якого визначається числом одиничних коренів (d). Якщо порядок $d = 1$, то тренд лінійний, і процес називається інтегрованим першого порядку. Якщо $d = 2$, то тренд квадратичний, а процес – інтегрований другого порядку і т.д. Якщо в одній моделі розглядаються інтегровані процеси з трендами однакового порядку, то такі процеси можуть бути коінтегрованими.

Загалом в теорії часових рядів розрізняють два типи трендів – детермінований та стохастичний. Детермінований тренд вказує на довгострокові зміни процесів, він не містить випадкову компоненту і його коефіцієнти не змінюються протягом тривалого проміжку часу. Прикладами детермінованих трендів є: лінійний, квадратичний, параболічний, логарифмічний тренди. Стохастичний тренд містить випадкову складову і його коефіцієнти швидко змінюються у часі. Отже, детермінований тренд характеризує закономірну динаміку розвитку процесу, а стохастичний тренд містить випадкові коливання (шуми) в процесі. Більшість часових рядів, які описують економічні процеси, містять стохастичні тренди і це обов'язково необхідно брати до уваги при побудові моделей (особливо регресійних).

Гетероскедастичний процес – нестаціонарний процес, в якому на інтервалі часу дослідження змінюється дисперсія (рис. 1.3).

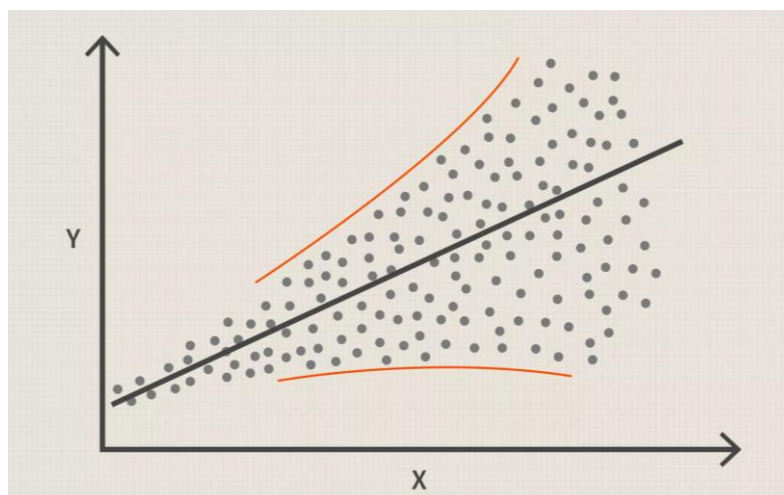


Рисунок 1.3 – Графік з випадковими даними, що показують гетероскедастичність процесу [8]

Для попереднього виявлення нелінійності та гетероскедастичності процесу можна скористатись візуальним аналізом даних. Це дозволяє оперативно виявити наявність ділянок з лінійним або нелінійним трендом, виявити наближено наявність гетероскедастичності та імпульсів, які можуть впливати на якість моделі. Також за допомогою візуального аналізу можна наближено оцінити тип розподілу за гістограмою та описовими статистичними параметрами.

1.3 Існуючі підходи до моделювання і прогнозування нелінійних нестационарних процесів та огляд результатів досліджень

Реальні фінансово-економічні, екологічні, соціальні та інші системи зазвичай демонструють нелінійну і нестационарну поведінку. Більшість методів прогнозування, які описано в літературі, припускають лінійність та/або стаціонарність системи [9] або ж розглядають прості форми нестационарності, такі як чітко визначені тенденції і варіації перших двох моментів, прості форми кусково-стаціонарних режимів [10].

На рисунку (Рис. 1.4) зображена класифікація різних методів для прогнозування різних даних часових рядів.

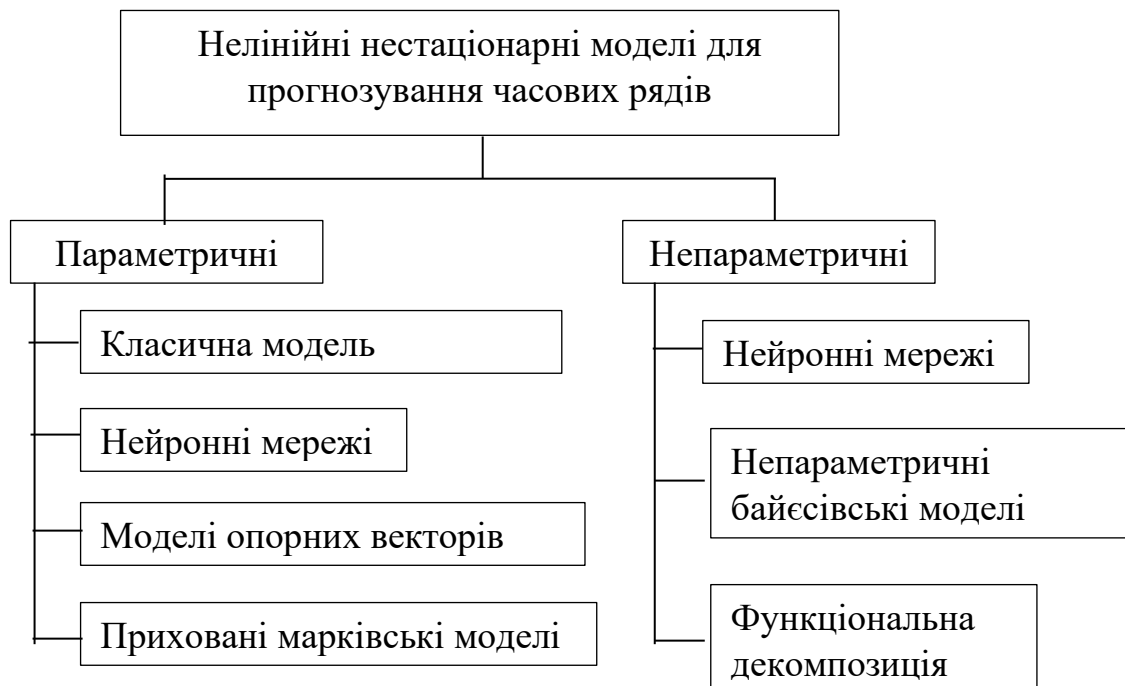


Рисунок 1.4 – Класифікація моделей для прогнозування нелінійних нестационарних процесів [1]

1.3.1 Параметричні моделі

Моделі для параметричного прогнозування визначають явну форму функції із скінченним числом параметрів для опису взаємозв'язків між вхідними даними, що складаються з внутрішніх та зовнішніх змінних і їх авторегресійних членів, і вихідними даними, що містять майбутні значення внутрішньої (основної) змінної [4].

Класична модель авторегресії

Моделі авторегресії (АР) найбільш широко вивчаються через їхню гнучкість для моделювання багатьох стаціонарних та деяких нестационарних процесів. До них відноситься модель авторегресії з ковзним середнім (АРКС),

яка передбачає лінійний зв'язок між запізнілими (тобто, минулими значеннями) змінних. Однак ці лінійні моделі є лише досить грубим наближенням до реальних складних систем і зазвичай не можуть з високою точністю передбачити еволюцію нелінійних і нестационарних процесів. Продуктивність моделі АРКС часто значно погіршується при наявності тимчасових тенденцій (явищ) і сезонних особливостей [4].

Такі моделі, як ARIMA використовуються для усунення/зменшення нестационарності першого порядку. Однак диференціювання, як правило, підсилює високочастотні шуми в часових рядах, і для визначення фактичного порядку моделі ARIMA часто потрібно докласти значних зусиль. Крім того, моделі ARIMA здебільшого обмежуються фіксацією нестационарності першого порядку [4].

Engle (1982) представив модель авторегресії з умовною гетероскедастичністю (ARCH) для фіксації нестационарності другого порядку [11]. Узагальнена модель авторегресії з умовною гетероскедастичністю (GARCH) була описана пізніше у роботі Bollerslev (1986) [12].

Для того щоб включити нелінійність як частину структури АРКС (ARMA), Tong [13] запропонував моделі порогової авторегресії (TAR), які передбачають кускову лінійність (тобто різні лінійні форми в різних областях простору). Моделі TAR успішно застосовуються для короткострокового прогнозування часових рядів в економіці та інших галузях [14].

McAleer і Medeiros [15] досліджували гетерогенні моделі авторегресії (AR) з багаторежимним плавним переходом для прогнозування фінансових часових рядів. Ця модель може апроксимувати мінливість і поведінку довгострокової пам'яті, а також асиметрію знаку та розміру, яка притаманна фінансовим часовим рядам.

Lundbergh [16] запропонував модель для одночасного розгляду перехідних режимів та коефіцієнтів, що змінюються в часі; тобто коефіцієнти AR змінюються як залежно від часу, так і від вхідної змінної, яка плавно перемикається між двома режимами. Однак на практиці багато параметрів мають

бути оптимізовані для побудови узгодженої STAR (Smooth Transition AR, модель авто регресії з плавним переходом) моделі, що змінюється з часом. Lineesh і John [17] вивчали трендові і порогові моделі AR для нелінійних і нестационарних часових рядів.

Класичні моделі для регресійного прогнозування, засновані на багатовимірних часових рядах, такі як Vector AR (VAR), також вивчалися з метою врахування впливу екзогенних змінних. Однак нестационарна еволюція таких змінних, яка може постійно відхилятися від попередніх станів, створює додаткові проблеми для довгострокового прогнозування цільових змінних.

Нейронні мережі

Нейронні мережі (НМ, англ. – NN) та інші нелінійні моделі на основі операторів використовуються для прогнозування нелінійних часових рядів у виробничих системах, фінансах, інформатиці, охорони здоров'я, енергетичних та в інших сферах [18], [19], [20]. Ці моделі не потребують попередніх припущень про форму нелінійності і є універсальними наближеннями [21]; тобто вони можуть апроксимувати будь-яку неперервну функцію з довільною точністю. Нещодавній огляд моделей НМ для прогнозування часових рядів був представлений у роботі Zhang [22].

Моделі нейронних мереж з прямим зв'язком (FNN), що параметризовані за допомогою алгоритму зворотного поширення, використовуються для прогнозування нелінійних часових рядів [23], [24]. Відомо, що вони кращі, ніж традиційні статистичні методи, такі як регресія і підходи Бокса-Дженкінса у функціональному наближенні, але вони припускають, що динаміка, яка лежить в основі часових рядів, не залежить від часу. Моделі FNN з рекурентними зворотними зв'язками також використовувались для прогнозування часових рядів у роботі De Groot [25]. Такі моделі рекурентних НМ (RNN) дозволяють прогнозувати нелінійні часові ряди, що виникають в різних областях [26], [27]. Menezes і Barreto [28] побудували рекурентну мережеву структуру нелінійних

моделей AP (AR) з екзогенними вхідними даними для багатоетапного прогнозування хаотичних часових рядів.

Різні типи НМ моделей з радіальною базисною функцією (RBF), такі як моделі, що використовують динамічну регуляризацию [29], навчання методом найменших квадратів [30] і рекурсію [31], були досліджені з метою виявлення різних форм тенденцій і нестабільності в часових рядах.

Barreto [32] розглянув підходи до прогнозування часових рядів з використанням SOM (карта, що самоорганізовується, англ. – Self-Organizing Map) НМ моделей.

Отже, нейронні моделі можуть обробляти довгі (великої потужності) нелінійні часові ряди, де основні фізичні взаємозв'язки невідомі [33].

Моделі опорних векторів (SVM)

Методи прогнозування на основі SVM використовують клас узагальнених регресійних моделей, таких як регресія опорних векторів (SVR) і машини опорних векторів найменших квадратів (LS-SVM; [34]), які параметризуються з використанням методів випуклого квадратичного програмування [35].

Mukherjee та ін. [36] (1977) досліджували застосування моделі SVM для прогнозування хаотичних часових рядів. Вони показали, що SVM мають більш високу точність прогнозування, ніж НМ моделі, і використовують менше параметрів.

Lau і Wu [37] розглянули предиктори на основі найменших квадратів (LS) і RBF (Radial Basis Function) і розробили локальну SVM (визначену в зміненому просторі станів) для прогнозування хаотичних часових рядів. Їх дослідження показали, що локальні моделі SVM можуть забезпечити більш високу точність для довгострокового прогнозування в порівнянні з локальними поліноміальними предикторами на основі LS і RBF.

Khemchandani та ін. [38] досліджували нечіткі моделі SVR для прогнозування нестационарних часових рядів, виходячи з основного припущення, що останні вибірки даних надають більш актуальну інформацію

для прогнозування. Такий підхід знижує надмірну підгонку і обчислювальні витрати, властиві традиційному SVR.

Приховані марківські моделі

Більшість розглянутих вище моделей включають пакетну обробку, при якій модель адаптується та періодично оновлюється з використанням історичних даних. Однак непомірно великі обчислювальні затримки, вимоги до пам'яті та великі розміри даних перешкоджають їх застосуванню до багатьох реальних проблем, особливо для онлайн моніторингу процесів. Для подолання цього обмеження досліджуються різні послідовні (також відомі як онлайн або рекурсивні) моделі прогнозування, такі як приховані марківські моделі (англ. НММ; [39]). НММ – це клас моделей, в якому часовий ряд розглядається як функція, що лежить в основі неспостережуваного вектору станів.

Одним з обмежень НММ є те, що їх ефективність прогнозування, як правило, особливо чутлива до порядку марківських символів, що використовуються для представлення стану. Крім того, параметрична форма моделей НММ може бути громіздкою, а то й неадекватною, для відображення еволюції сильно нестационарних часових рядів. Отже, НММ привабливі для моделювання еволюції часових рядів, що генеруються з дискретного простору станів, де перемикання між кінцевим станом слідує певній закономірності (наприклад, коливання цін на акції [40]).

1.3.2 Непараметричні моделі

Хоча параметричні моделі можуть забезпечувати точні прогнози, коли моделі задані правильно, вони мають тенденцію ставати дуже неоптимальними, коли динаміка, що лежить в основі, невідома або невизначена. Крім того,

проблема зміщення моделі зберігається, оскільки динаміка більшості реальних складних систем за своєю суттю нелінійна і нестационарна [4].

У той же час непараметричні моделі можуть забезпечити повне представлення динаміки на основі спостережуваних даних, не передбачати ніяких структурних припущень і спростити моделювання. Отже, точність моделювання і прогнозування для нелінійних і нестационарних часових рядів поліпшується. Однак, у порівнянні з параметричними моделями, непараметричні моделі зазвичай вимагають великих наборів даних, з яких можна ефективно взяти інформацію про взаємозв'язки між даними.

Підходи на основі найближчого сусідства

Підходи на основі найближчого сусідства передбачають майбутні значення шляхом повторної вибірки історичних спостережень з основним припущенням, що майбутня поведінка змінюється плавно; тобто спостереження, аналогічні цільовим, швидше за все, будуть мати аналогічні результати. Ці моделі привабливі для прогнозування динаміки складних систем своєю простотою і достатньою точністю [41]. Namid і ін. [42] досліджували різні методи, засновані на сусідстві, для прогнозування хаотичних часових рядів; наприклад, моделі нульового порядку (один найближчий сусід), k найближчих сусідів (кілька сусідів) і зважене наближення відстані (зважене середнє значення декількох сусідів).

Байєсівські непараметричні моделі

Байєсівське моделювання – це в основному процес об'єднання апріорної інформації з експериментальною (статистичною) для отримання апостеріорного висновку; тобто для оцінювання умовного розподілу прихованої моделі або параметрів з урахуванням характеристик часового ряду, що спостерігається [43].

На відміну від інших байєсівських методів, байєсівські непараметричні моделі припускають, що прихована структура тут розвивається разом з даними. Іншими словами, байєсівські непараметричні моделі спрямовані на пошук єдиної

моделі з нескінченної кількості можливих, складність якої адаптується відповідно до даних (наприклад, кількість параметрів для оцінювання). Серед байєсівських непараметричних моделей такі: моделі GP (гаусівський процес, англ. – Gaussian Process) були найбільш широко вивчені для прогнозування часових рядів [44]. Останнім часом також було вивчено багатоетапне прогнозування часових рядів за допомогою моделей GP. Girard та ін. [45] вивчали модель GP для багатоступінчастого прогнозування часових рядів із вхідною невизначеністю, де вхідні дані для прогнозування наступного кроку були отримані шляхом формування вибірки вихідного розподілу, заданого в термінах середнього і дисперсії, а також оцінених на попередньому часовому кроці гаусівського процесу.

Моделі функціональної декомпозиції

В останні роки особлива увага приділяється моделям функціональної декомпозиції [46] для прогнозування часових рядів. Для побудови базису декомпозиції використовуються локальні характеристики часового ряду, включаючи інтервал між послідовними локальними екстремумами (максимумами і мінімумами) в різних часових масштабах. Ці моделі можуть використовуватися для фіксації дрейфів і нелінійних режимів будь-яких нелінійних нестационарних процесів. Більшість методів прогнозування в цій категорії мають гібридний підхід, оскільки вони використовують модель декомпозиції для представлення часових рядів та інший параметричний або непараметричний метод для прогнозування розвитку окремих компонентів. Функціональна декомпозиція часто зменшує або розподіляє за різними функціональними компонентами ефекти нестационарності (наприклад, тренд і сезонність) в часових рядах.

Мультиплікативна декомпозиція також може використовуватися для виявлення тенденцій сезонних коливань; тобто часовий ряд можна розглядати як мультиплікацію сезонних і несезонних складових. Якщо сезонні коливання

відносно постійні з наявністю тренду, зазвичай краще використовувати адитивну декомпозицію [47].

Отже, аналіз літературних джерел свідчить, що останнім часом зріс інтерес до непараметричних моделей для прогнозування нелінійних і нестационарних часових рядів. Адаптація принципів нелінійних динамічних систем до непараметричних підходів моделювання може бути засобом для подальшого покращення якості прогнозування в умовах періодичних змін та інших кусково-стационарних умовах.

1.4 Деякі сучасні програми для моделювання і прогнозування ННП

Сучасні завдання статистичного аналізу даних в різноманітних сферах діяльності людини сприяють розвитку ринку комп'ютерних програм для статистичної обробки даних. На сьогоднішній день доступність програм статистичних пакетів значно спростила можливість легко застосовувати будь-яку статистичну модель, засновану на типах даних, таких як часові ряди, панельні дані та інші. [48] Використання статистичних програмних продуктів є невід'ємною частиною проведення економетричного аналізу, оскільки це спрощує процес побудови моделей, прийняття рішень щодо специфікації та ідентифікації моделі, вибору методу оцінки параметрів моделі, інтерпретації результатів, отримання прогнозних оцінок, тощо. Програми постійно удосконалюються: прискорюється робота з даними, покращується представлення результатів аналізу даних, стає зручнішим інтерфейс, збільшується число вбудованих в програму статистичних процедур, засобів обробки даних і ін.

За функціональними можливостями статистичні пакети класифікують як:

- універсальні пакети (EViews, R, STATISTICA, STATA);
- професійні пакети (SAS, BMDP);

– спеціалізовані пакети (BioStat, DATASCOPE, MESOSAUR).

При виборі належного (раціонального) статистичного пакету варто звертати увагу на такі важливі параметри як: зручність використання, вбудована кількість методів для аналізу, об'єми даних, які програма здатна обробляти та швидкість обробки, наявність методичного забезпечення, висока якість роботи і доступність.

Розглянемо найбільш поширені з цих статистичних пакетів для моделювання і прогнозування нелінійних нестационарних процесів, а саме: Statistica, Eviews, SAS, Python (бібліотека SciPy).

1.4.1 Statistica

Statistica розроблена для статистичного аналізу даних як одновимірного, так і багатовимірного характеру. Система має модульну структуру, де кожний модуль є повноцінною програмою, що вирішує конкретну групу завдань. Програма надає потужні та зручні у використанні інструменти для отримання даних (data mining), для статистичного та графічного аналізу даних (data analysis), для управління даними (data management), для прогнозування та візуалізації даних (data visualization), для створення власних користувацьких додатків та інші. Інтерфейс Statistica є зручним та інтуїтивно зрозумілим, легко налаштовується відповідно до будь-яких завдань і аналогічний інтерфейсу стандартних Windows додатків.

Також до переваг даної програми належать:

- Можливість роботи з файлами необмежених розмірів та швидкодія при роботі з великими об'ємами даних, потужні обчислювальні можливості, достатні для регулярного створення запитів до баз даних та комплексного управління даними.
- Висока точність статистичних розрахунків.

- Наявність потужного графічного модуля з інструментами для ефективної візуалізації даних, редагування графічних матеріалів та проведення графічного аналізу. Представлено близько 10 тисяч типів графіків, матриць і пікрограм, які можна редагувати та інтерактивно налаштовувати (обертання, масштабування, накладання графіків, накладання спеціальних ефектів і т.д.).
- Широкі можливості інтеграції та сумісності, що дають змогу впроваджувати нові модулі в існуючі системи та будувати на основі Statistica інтелектуальні системи прийняття рішень.
- Простий імпорт/експорт даних, оскільки Statistica взаємодіє з будь-якими реляційними базами даних (Oracle, MS SQL Server, Informix, Access та ін.), програмами (MS Excel, MS Access, Foxpro, Paradox, dbase, CSV, SPSS) та сховищами бізнес-інформації (SAP Business Warehouse).
- Автоматизація будь-яких функцій і процедур за допомогою макросів. Написані мовою програмування Statistica Visual Basic макроси дають змогу створювати користувацькі програми і надбудови для автоматизації щотижневих або тривалих процедур.
- Підтримка Web-технологій – введення і виведення даних через Web-браузер, направлення результатів на Web-сервер.

Недоліками системи Statistica є такі:

- Відсутність реалізації деяких важливих тестів часових рядів (тест на коінтеграцію, VAR, GARCH).
- Недостатньо опрацьований блок методів, націлених на економетричне моделювання.
- Висока вартість програмного продукту (понад 3000 \$).
- Працює лише в операційній системі Windows.

1.4.2 EViews

Eviews – один з найпотужніших статистичних пакетів, призначений для загального статистичного аналізу та економетричного аналізу даних часових рядів, прогнозування та моделювання панельних даних, побудови регресійних моделей. Eviews поєднує в собі електронні таблиці та реляційні бази даних із традиційними завданнями статистичного програмного забезпечення і використовує графічний інтерфейс. Пакет використовується для вирішення наступних завдань: аналіз наукової інформації, фінансовий аналіз, макроекономічне прогнозування, моделювання економічних процесів, прогнозування станів ринків [49].

Перевагами Eviews є такі:

- Інтуїтивно зрозумілий і простий у користуванні об'єктно-орієнтований інтерфейс користувача. Можливість одночасно відображати кілька об'єктів, кожен у своєму вікні. Підтримка декількох вікон спрощує паралельне порівняння графіків, перевірку гіпотез, оцінку рівнянь або прогнозів моделей, розроблених за альтернативними припущеннями.
- Забезпечує зручні візуальні способи введення рядів даних вручну (з клавіатури) або з файлів, створення нових рядів даних з вже існуючих, відображення та проведення статистичного аналізу взаємозв'язків між даними.
- Підтримка багатьох форматів даних та можливість імпортувати їх з різних баз даних, включаючи Excel, SAS, SPSS та STATA, зазначені у форматах .xls, .sas, .sav та .dta, також можна імпортувати текст (.txt), двійкові файли, файли HTML тощо.
- Функціональні переваги. Eviews пропонує такі функції, як інтерполяція даних та параметри частотного фільтра, за допомогою яких можна згенерувати відсутні дані, що є корисним для аналізу

часових рядів. База даних часто вимагає довгострокових розрахунків дисперсії та коваріації, в межах яких помилки розподіляються випадковим чином і вказують на певний тип послідовності. Eviews дає можливість розрахувати симетричні або односторонні довгострокові дисперсії з використанням непараметричного ядра.

- Доступність (безкоштовний) [49].

До недоліків Eviews можна віднести такі:

- Eviews обмежений матрицею ряду спостережень і змінних, тому не підходить для проведення аналізу панельних даних з довгостроковими вибірками даних.
- Відсутня можливість реалізації власних алгоритмів.
- Слабкі можливості візуалізації (в порівнянні з STATISTICA).
- Абсолютно інші принципи роботи в порівнянні з STATISTICA або Excel.

1.4.3 SAS

SAS – потужна і складна система для статистичної обробки даних. Програмне забезпечення складається з різних модулів, деякі з яких мають абсолютно різні принципи роботи, відповідно необхідно мати певні знання і навички. Всі програми складаються з двох частин: перша частина – етап даних (створення нових та зміна існуючих наборів даних) та друга частина – етап процесу (виконання аналізу на основі вибраних наборів даних).

Перевагами SAS є такі [50]:

- Розпізнавання будь-якого типу даних і можливість отримувати доступ до даних із будь-якого програмного забезпечення та будь-якого формату.
- Гнучкий інтерфейс обміну даними.

- Наявність інструментів для роботи з кластерами (розподіленими системами).
- Велика швидкість розрахунків на великих масивах даних.
- Надає повні, всебічні інструменти для аналізу даних, такі як: багатовимірні методи, описова статистика, лінійне програмування, регресійний аналіз, байєсівський аналіз, просторовий аналіз та інші. Підтримка ARIMA, GARCH, VAR.
- Швидка інтеграція нових статистичних методів.
- Наявність детальної документації та професійної сервісної підтримки.
- Доступні численні модулі та інтерфейси, а також окреме програмне забезпечення для бізнес-аналітики.

Недоліками SAS є:

- Для написання скриптів необхідно знати мову програмування SAS Macro Language.
- Висока вартість ліцензії. Порівняно з іншими статистичними пакетами, SAS є одним із найдорожчих рішень, тому більше орієнтоване для використання великими компаніями та організаціями.
- Складність освоєння програми та вимоги до математичних знань користувача.
- Частково застарілий користувацький інтерфейс.
- Об'ємність ПЗ та високі вимоги до апаратної частини.

1.4.4 Python (бібліотека SciPy)

Python – повністю функціональна, відкрита, інтерпретована мова програмування. Особливо добре підходить для глибокого та машинного

навчання, а також в якості статистичного програмного забезпечення завдяки використанню бібліотек, які легко встановлюються.

Бібліотека SciPy містить необхідні модулі для проведення статистичного аналізу даних: scikit-learn (наявні різні моделі машинного навчання: алгоритми класифікації, кластеризації, регресії, тощо) [51] і statmodels (можливість використання статистичних моделей та тестів: регресія, побудова графіків, узагальнена лінійна модель, аналіз часових рядів, модель авторегресії ковзного середнього, векторна авторегресія, непараметрична статистика, емпірична ймовірність і т.д.)[52].

Переваги Python (бібліотека SciPy):

- Потужна і функціональна мова програмування.
- Легко автоматизується (наприклад, за допомогою скриптів або веб-сервера).
- Повністю інтегрується (Git, teradata, PySpark, Hadoop).
- Доступність (безкоштовна).
- Проста у створенні та зрозуміла візуалізація, наявність готових засобів для візуалізації даних.

Недоліки Python (бібліотека SciPy):

- Доступні не всі статистичні методи.
- Необхідні навички програмування.

Висновки до розділу 1 та постановка задачі дослідження

Часові ряди складних систем реального світу демонструють наявність різноманітних нелінійних нестационарних динамічних процесів. Їх прогнозування необхідне для коректного планування розвитку та вдосконалення способів управління і контролю складними реальними системи – від

традиційного виявлення – діагностики – пом'якшення наслідків до більш попереджувального прогнозування – прогнозу – запобігання.

Зі збільшенням багатовимірних даних стосовно розвитку процесів існує необхідність у розробці моделей для ефективного знаходження прихованих закономірностей у складних системах, зменшення розмірності даних, визначення динамічних взаємозв'язків та використання компактного набору факторів для точного прогнозування.

В цьому розділі розглянуто основні типи нестационарних процесів та їхні характерні особливості. Зроблено порівняльний аналіз різних програмних продуктів для статистичної обробки даних, визначено основні переваги та недоліки кожного з них. Здійснено огляд існуючих підходів до моделювання і прогнозування нелінійних нестационарних процесів: параметричних (класична авторегресія, нейронні мережі, SVM, HMM) і непараметричних (ближнє сусідство, байєсівські моделі, функціональна декомпозиція), також розглянуто останні дослідження та розробки в області прогнозування процесів.

Постановка задачі дослідження:

- Виконати огляд існуючих типів нелінійних нестационарних процесів (ННП).
- Виконати аналіз сучасних методів, які використовуються для моделювання і прогнозування ННП в економіці та фінансах.
- Виконати огляд деяких результатів відомих досліджень.
- Вибрати методи для моделювання і прогнозування ННП в економіці та фінансах.
- Вибрати методику для моделювання ННП.
- Розробити програмний продукт для моделювання і прогнозування ННП; виконати обчислювальні експерименти з метою моделювання.
- Виконати порівняльний аналіз отриманих результатів і зробити висновки.
- Виробити рекомендації щодо моделювання ННП.

РОЗДІЛ 2 ВИБРАНІ МЕТОДИ І ПІДХОДИ ДО МОДЕЛЮВАННЯ І ПРОГНОЗУВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ФІНАНСАХ

2.1 Регресійний підхід

У регресійному аналізі використовуються регресії для кількісної оцінки взаємозв'язку між однією випадковою змінною та іншими випадковими змінними. При цьому, перша змінна називається залежною, а решта – незалежними. Вибір або призначення залежних і незалежних змінних є довільним (умовним) і здійснюється в залежності від задачі, яка розв'язується. Незалежні змінні називаються факторами, регресорами або предикторами, а залежна змінна – результативною ознакою або реакцією [53].

Якщо число предикторів дорівнює 1, регресію називають простою або однофакторною, якщо число предикторів більше 1 – множинною або багатофакторною. У загальному випадку регресійну модель можна записати в такий спосіб:

$$Y = f(x_1, x_2, \dots, x_n), \quad (2.1)$$

де y – залежна змінна (відгук), x_i ($i = 1, \dots, n$) – предиктори (фактори), n – число предикторів.

За допомогою регресійного аналізу можна вирішувати такі завдання:

- Зменшення розмірності простору змінних, що аналізуються (факторного простору) за рахунок заміни частини факторів однієї змінної – реакції.
- Проаналізувати кількісне вимірювання ефекту кожного фактору, множинна регресія дозволяє задати питання про те, «що є кращим предиктором для ...». При цьому, стає більш ясним вплив окремих

факторів на залежну змінну, і можна краще зрозуміти природу досліджуваного явища.

- Обчислення прогнозних значень змінної при певних значеннях факторів, тобто регресійний аналіз, створює базу для обчислювального експерименту з метою отримання відповідей на питання типу «Що буде, якщо ...».
- У регресійному аналізі в більш явній формі виступає причинно-наслідковий механізм. Прогноз при цьому краще піддається змістовній інтерпретації.

2.2 Метод групового урахування аргументів (МГУА)

Метод запропоновано наприкінці 60-х років ХХ століття академіком О.Г. Івахненком (Інститут кібернетики НАН України). Він ґрунтується на ідеях самоорганізації і механізмах живої природи – схрещуванні (гібридизації) і селекції (доборі). МГУА використовують для моделювання складних систем, короткострокового та довгострокового прогнозування, виявлення фізичних закономірностей, ідентифікації та апроксимації багатофакторних систем, розпізнавання образів та кластеризації вибірки даних. В результаті застосування тільки цього індуктивного методу до неточних, зашумлених та/або коротких вибірок даних може бути знайдена оптимальна математична модель [54].

В основі методу лежить система опорних функцій, правила ускладнення моделі, критерій селекції та регуляризація згідно зовнішнім критеріям. Спочатку проводиться генерація моделей-претендентів, потім селекція згідно зовнішнім критеріям та відсів моделей, що не пройшли селекцію. Алгоритм самоорганізації можна зобразити у такому вигляді:

- попередня обробка спостережень з урахуванням системи обраних опорних функцій (скорочується кількість претендентів);

- генерування множини моделей-претендентів;
- обчислення критеріїв селекції та пошук моделі оптимальної складності.

Для задачі прогнозування доцільно зменшити точність обрахунку оцінок коефіцієнтів рівняння регресії, проте надати йому більшу регулярність. Тому метою є не мінімізація похибок на вже відомих вузлах інтерполяції, а мінімізація похибок на нових точках [55].

У поліноміальних алгоритмах методу групового урахування аргументу послідовність вихідних даних ділиться на перевірочну та навчальну, де M – кількість вузлів інтерполяції, а m — кількість членів повного поліному регресії. При $m > M$ розв'язок можливо отримати тільки за допомогою МГУА.

Нехай повний опис об'єкту задається залежністю $y = \varphi(x_1, x_2, \dots, x_n)$. Замінімо цей вираз декількома рядами часткових описів. Перший ряд селекції:

$$y_1 = f(x_1, x_2), y_2 = f(x_1, x_3), \dots, y_s = f(x_{n-1}, x_n), s = C_n^2. \quad (2.2)$$

Другий ряд селекції:

$$z_1 = f(y_1, y_2), z_2 = f(y_1, y_3), \dots, z_p = f(z_{n-1}, z_n), p = C_s^2 \dots \quad (2.3)$$

Алгоритми МГУА розрізняються за видом функції часткового опису $f(\cdot)$.

До основних видів часткових описів належать:

лінійний частковий опис —

$$f(x_i, x_j) = A_0 + A_1 x_i + A_2 x_j \quad (2.4)$$

неповний квадратичний частковий опис —

$$f(x_i, x_j) = A_0 + A_1x_i + A_2x_j + A_3x_ix_j \quad (2.5)$$

квадратичний частковий опис —

$$f(x_i, x_j) = A_0 + A_1x_i + A_2x_j + A_3x_ix_j + A_4x_i^2 + A_5x_j^2, \quad (2.6)$$

де A_i – нечітке число.

За допомогою порогових відборів в наступний ряд селекції передається тільки певна кількість регулярних або незміщених змінних. Ця кількість однакова на всіх рядах селекції, та дорівнює F .

Правило зупинки селекції: ряди селекції нарощуються до тих пір, доки критерій незміщеності розв'язків зменшується. Для запобігання перенавчання, селекцію при досягненні мінімуму необхідно зупинити [54].

2.3 Методика моделювання нелінійних нестационарних процесів

2.3.1 Методика побудови математичних моделей часових рядів

Для побудови моделей, для яких складно поставити експеримент та отримати інформативні експериментальні дані в достатньому об'ємі можна використати методику, запропоновану Боксом і Дженкінсом.

Алгоритм побудови математичної моделі процесу містить такі етапи [56]:

- а) Проведення аналізу процесу, для якого будується модель, на основі літературних джерел, експертних оцінок, візуального дослідження вимірів вхідних і вихідних змінних, представлених часовими рядами, та іншої доступної інформації.
- б) Попередня обробка експериментальних даних з метою приведення їх до найбільш зручного вигляду для оцінювання параметрів моделі.

- в) Перевірка часових рядів на можливу наявність нестационарності і нелінійності з використанням статистичних тестів.
- г) Оцінювання структур моделей-кандидатів:
- 1) Обчислення та аналіз кореляційної матриці для часових рядів залежної та незалежних змінних з метою визначення тих екзогенних змінних, які необхідно включити в модель.
 - 2) Обчислення автокореляційної (АКФ) та часткової автокореляційної функції (ЧАКФ) залежної змінної з метою оцінювання порядку авторегресійної частини моделі.
 - 3) Розрахунок можливих нелінійних кореляцій між змінними.
 - 4) Оцінювання характеристик інших елементів структури математичної моделі.
- д) Вибір методу (або методів) оцінювання параметрів математичних моделей вибраних структур. Найчастіше це метод найменших квадратів (МНК), метод максимальної правдоподібності (ММП) та їхні модифікації (рекурсивні та нелінійні).
- е) Вибір кращої з оцінених моделей-кандидатів за допомогою множини статистичних критеріїв адекватності (якості) моделі. Застосування моделі для прогнозування і остаточне встановлення придатності моделі.

Аналіз процесу

Ігнорування цього етапу призводить до неможливості побудови моделі високого ступеня адекватності та її придатності для розв'язання задач. Аналіз процесу включає наступні завдання[56]:

- визначення вимірності процесу (числа входів і виходів процесу);
- встановлення логічних зв'язків між змінними та можливостей коректного об'єднання в одному математичному виразі;
- визначення кількості зовнішніх збурень та їх типу (детерміноване чи стохастичне);
- з'ясування можливості декомпозиції процесу – розкладання процесу на окремі, більш прості підпроцеси;

- розмежування рівнів процесу, якщо він має ієрархічну структуру (верхній та нижній рівень функціонування), визначення функції кожного з них і встановлення типів зв'язків між ними;
- аналіз спеціальної літератури з метою виявлення особливостей функціонування процесу, відомих законів та закономірностей його протікання, пошуку існуючих моделей процесу та досвіду його теоретичного чи експериментального дослідження;
- встановлення недоліків та переваг розроблених моделей досліджуваного процесу, а також можливості подальшого використання (модифікації). Аналіз і використання існуючих моделей дає можливість суттєво скоротити час та інші витрати на побудову та використання моделі.

Отриману інформацію використовують для попереднього оцінювання структури моделі або декількох моделей-кандидатів, параметри яких оцінюють за допомогою експериментальних даних. У процесі виконання аналізу функціонування досліджуваного процесу варто використовувати та порівнювати інформацію з різних джерел. Особливо це стосується фінансово-економічних процесів, де інформація може надходити з істотними протиріччями, пропусками і похибками.

Попередня обробка (підготовка) експериментальних даних

Процес обробки експериментальних даних, складається з наступних дій:

- нормування та візуальна перевірка даних;
- корегування даних за необхідності, а саме заповнення пропусків та зменшення екстремальних значень, що виходять за межі діапазону значень змінних;
- формування перших або різниць вищих порядків, які необхідні для аналізу відповідних складових часового ряду.

Під нормуванням даних мається на увазі приведення їх до зручного діапазону, наприклад, від 0 до 1, від -1 до 1 і ін. Поширеним методом нормування даних є їх логарифмування з наступним формуванням додаткових часових рядів з перших чи других різниць. Перші різниці є наближеним дискретним аналогом

першої похідної, а другі різниці – другої похідної. Також для того, щоб працювати не з повними значеннями, а їх відхиленнями, з кожного значення в ряді віднімається його середнє. Залежно від випадку, застосовують різні способи для нормування даних.

Виявлення і обробка екстремальних та аномальних значень є важливою складовою на етапі попередньої обробки даних. Аномальні дані – дані, які виникли внаслідок впливу значних похибок вимірів або через некоректний збір статистичних даних. Якщо встановлюється наявність аномальних даних, тоді їх видаляють з розподілу [56].

Екстремальні значення – це дані, які зібрано коректно і які містять раптові стрибкоподібні зміни в процесі. Існують різні методи дослідження екстремальних значень. У випадку якщо потрібно лише виявити такі значення (наприклад, щоб знайти причини, які приводять до появи цих екстремальних значень), достатньо використати критерії для виявлення таких значень. А якщо слід виявити і виключити екстремальні значення (наприклад, щоб покращити оцінки статистичних параметрів і моделей), тоді виникає задача – як правильно виконати обробку даних. Таким чином, за допомогою критерію для визначення екстремальних значень, можна визначити величину зміщення оцінок параметрів.

Критерії аналізу екстремальних значень застосовують з метою:

- вирівнювання спостереження перед аналізом (зменшити великі значення);
- перевірки наявності в даних аномальних значень, якщо вони є – тоді варто переглянути процедуру збору даних;
- виокремлення спостережень, які є цікавими з точки зору аномальності та їх математичний опис.

Класичний підхід до виявлення аномальних спостережень полягає в тому, що вибіркові спостереження розглядають як випадкові, нормально розподілені величини. При цьому для виявлення екстремальних значень створюється статистика (статистичний тест), яка є чутливою до різких відхилень такого типу. Необхідно встановити розподіл цієї статистики при нульовій гіпотезі, що всі

спостереження належать нормальній сукупності, а потім відхилити цю гіпотезу, якщо виявиться, що обчислена статистика їй протирічить [56].

Аналіз наявності нелінійності та нестационарності

Однією з важливих проблем при визначенні структури моделі є з'ясування наявності нелінійностей в досліджуваному процесі та їх типу. Щоб вирішити цю проблему використовують візуальний аналіз даних та формальні тести. Візуальний аналіз дозволяє ефективно виявити наявність ділянок з лінійним або нелінійним трендом, в певній мірі чи є гетероскедастичність та значні викиди (імпульси), які істотно впливають на якість моделі.

Також існує ряд формальних тестів для перевірки наявності нелінійності. Розглянемо один із тестів, що застосовується у випадку, коли можна набрати кілька груп (вибірок) спостережень для одного і того ж процесу:

$$\hat{F} = \frac{\frac{1}{m-2} \sum_{i=1}^m n_i (\bar{y}_i - \hat{y}_i)^2}{\frac{1}{n-m} \sum_{i=1}^m \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2}, \quad (2.7)$$

де \bar{y}_i – середнє значення для i -ї групи (вибірки або групи) даних; \hat{y}_i – середнє для лінійної апроксимації даних; m – число груп даних; n_i – число вимірів в i -й групі; n – загальне число вимірів. Фактично, наведена статистика представляє собою таке відношення:

$$\hat{F} = \frac{\text{Відхилення середніх значень від прямої регресії}}{\text{Відхилення значень } y(k) \text{ від групових середніх}} \quad (2.8)$$

У випадку якщо статистика \hat{F} з степенями свободи $\nu_1 = m - 2$ та $\nu_2 = n - m$ досягає або перевищує рівень значимості, то гіпотезу щодо лінійності потрібно відхилити. Недоліком цього підходу є те, що для його застосування необхідно мати декілька (не менше трьох) груп даних для одного і того ж

процесу, які можна отримати в результаті виконання повторних експериментів, але це не завжди можливо [56].

Також наявність нелінійності можна встановити за допомогою вибірових нелінійних кореляційних функцій (НКФ), які розраховані за вибірками експериментальних (статистичних) даних, наприклад, такої:

$$r_{yx^2}(s) = r_{y(k)x^2(k-s)} = \frac{1}{N} \frac{\sum_{k=s+1}^N \{[y(k) - \bar{y}][x(k-s) - \bar{x}]^2\}}{\sigma_y \sigma_x^2}, \quad s = 0, 1, 2, 3, \dots, (2.9)$$

Наприклад, якщо дискретна НКФ (2.4) містить значення, які статистично суттєво відрізняються від нуля, то процес містить квадратичну нелінійність відносно регресора x .

Наявність нелінійного детермінованого тренду в процесі можна визначити шляхом оцінювання наступного рівняння, яке представляє собою поліном порядку m відносно часу:

$$y(k) = a_0 + c_1 k + c_2 k^2 + \dots + c_m k^m, \quad (2.10)$$

Якщо хоча б один із коефіцієнтів c_i , $i = 1, \dots, m$ є статистично значимим, то гіпотеза щодо відсутності тренду відхиляється. Якщо тренд відносно швидко змінює свій напрям розвитку і його складно адекватно функціональне описати, то тоді використовують моделі випадкових трендів, які ґрунтуються на комбінаціях випадкових величин. Наявність нелінійності стосовно регресора $x(k)$ можна визначити за допомогою полінома:

$$y(k) = a_0 + c_1 x(k) + c_2 x^2(k) + \dots + c_m x^m(k) + \varepsilon(k) \quad (2.11)$$

Автоматично оцінює структуру математичної моделі метод групового врахування аргументів (МГУА). Цей метод застосовують для широкого класу процесів, в тому числі при моделюванні процесів різної природи з нелінійностями та нестационарностями. Подальшим розвитком даного методу є нечіткий МГУА, який ґрунтується на нечіткому представленні параметрів моделі, яка оцінюється [56].

Формування інших елементів структури моделі

На цьому етапі формально визначається структура моделей-кандидатів. Структура моделі складається з наступних елементів: порядку моделі (максимальний порядок рівнянь, які описують модель), розмірність (число рівнянь моделі), час затримки по входу та її реакція, нелінійності і їх тип, зовнішні збурення і їх тип (детерміновані або стохастичні, адитивні або мультиплікативні). Визначення складових та формування моделей-кандидатів дають можливість оцінити їхні коефіцієнти.

Оцінювання коефіцієнтів моделей-кандидатів

Для оцінювання коефіцієнтів (параметрів) рівняння застосовують принцип економії або збереження. Він означає, що кількість коефіцієнтів, що оцінюється, не повинна перевищувати їх необхідну кількість (наприклад, необхідність збереження в моделі основних статистичних характеристик процесу). Також варто пам'ятати, що чимало процесів, що досліджуються, є складними за поведінкою, тому їх потрібно апроксимувати за допомогою рівнянь, а не описувати до дрібниць.

Модель, яка оцінюється, повинна задовольняти принципу інверсії, тобто щоб за допомогою отриманого рівняння можна було б згенерувати початковий ряд, на основі якого оцінювались структура і коефіцієнти. Це означає, що хоча модель і спрощена, вона має співпадати з досліджуваним процесом за такими основними характеристиками як середнє, дисперсія та коваріація [56].

Найбільш поширеними методами оцінювання параметрів моделі є такі:

- метод найменших квадратів (МНК);
- метод максимальної правдоподібності (ММП);

- метод допоміжної (інструментальної) змінної (МДП);
- нелінійний метод найменших квадратів (НМНК) та їх рекурсивні версії (РМНК, РММП, РМДП);
- метод Монте-Карло для марковських ланцюгів (МКМЛ).

2.3.2 Множини критеріїв для аналізу якості даних, адекватності моделей та якості оцінок прогнозів

На цьому етапі проводиться аналіз якості моделі (наприклад, перевірка наявності кореляції між значеннями похибок, перевірка значимості параметрів, перевірка наявності нестационарності), що складається з наступних кроків:

- 1) Візуальне дослідження графіка похибок моделі

$$e(k) = y(k) - \hat{y}(k), \quad (2.12)$$

де $\hat{y}(k)$ – оцінка змінної, отримана за допомогою побудованої моделі.

На графіку не повинно бути значних викидів та довгих інтервалів, на яких похибка приймає великі значення (тобто довгих інтервалів суттєвої неадекватності). У випадку застосування рекурсивних методів оцінювання найбільші похибки будуть в перехідному процесі, коли інформаційна матриця ще не містить достатньо інформації про процес.

- 2) Похибки моделі не мають бути корельовані між собою.

Корельованість похибок визначається за допомогою статистики Дарбіна-Уотсона (DW) за формулою:

$$DW = 2 - 2\rho, \quad (2.13)$$

де $\rho = E[e(k)e(k-1)]/\sigma_e^2$ – коефіцієнт кореляції між сусідніми значеннями похибки; σ_e^2 – дисперсія послідовності похибок $\{e(k)\}$.

Якщо між похибками відсутня кореляція $DW = 2$ – це найкраще значення для параметру. Граничними значеннями для DW є 0 (при $\rho = 1$) та +4 (при $\rho = -1$) [56].

3) Перевірка на наявність нестационарності в процесі.

Для лінійної моделі 2-3 порядку оцінки параметрів повинні збігатися до сталого значення після 30-40 (не більше) ітерацій алгоритму оцінювання. Якщо кількість ітерацій значно перевищує вказані числа, то це свідчить про те, що процес може бути нестационарним.

4) Перевірка значимості параметрів моделі.

Для визначення значимості оцінки кожного коефіцієнта використовується статистика Стюдента або t – статистика (випадкова величина, що має t – розподіл), визначається за виразом:

$$t = \frac{\hat{a} - a^0}{SE_{\hat{a}}}, \quad (2.14)$$

де \hat{a} – оцінка коефіцієнта моделі; a^0 – нуль-гіпотеза (початкова гіпотеза) щодо цієї оцінки; $SE_{\hat{a}}$ – стандартна похибка оцінки коефіцієнта.

Статистична теорія перевірки гіпотез пропонує висувати нуль-гіпотезу, яка є протилежною бажаному результату, наприклад, що коефіцієнт значимий, тобто, $(H_0 : a^0 \neq 0)$ або незначимий $(H_0 : a^0 = 0)$. В даному випадку бажаним результатом є значимість коефіцієнтів математичної моделі. Отже, необхідно висувати нульову гіпотезу, що коефіцієнт незначимий. Це дає можливість правильно визначити значимості оцінок коефіцієнтів і спростити розрахунки [56].

Для встановлення значимості оцінки коефіцієнта потрібно знати довжину вибірки даних N (потужність вибірки); число ступенів свободи $f = N - n$, де n – число коефіцієнтів моделі, які оцінюються на основі ряду даних, і вибрати рівень значимості $\alpha = 1\%$ або $\alpha = 5\%$ або $\alpha = 10\%$ (для цих значень існують розраховані таблиці для критичних значень t – статистики).

Користуючись значеннями N , f і α , з таблиць для t – розподілу знаходять критичне значення t – статистики, тобто $t_{кр}$. Для перевірки правильності висунутої гіпотези розраховане значення t порівнюють з критичним $t_{кр}$. Якщо

$$-t_{кр} < t < t_{кр} \quad \text{або} \quad |t| < |t_{кр}|, \quad (2.15)$$

то нуль-гіпотеза щодо незначимості коефіцієнта приймається (його можна не враховувати в регресії). Звідси випливає, що чим більшим є значення t – статистики для оцінки коефіцієнта, тим імовірніше, що цей коефіцієнт є значимим.

5) Коефіцієнт множинної детермінації R^2 , який обчислюється так:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST}, \quad (2.16)$$

де $\text{var}(\hat{y})$ – дисперсія залежної змінної, оціненої за допомогою побудованої

моделі; $\text{var}(y)$ – дисперсія вимірів залежної змінної; $SSE = \sum_{k=1}^N [y(k) - \hat{y}(k)]^2$ –

сума квадратів похибок (залишків) моделі (sum of squared errors);

$SST = \sum_{k=1}^N [y(k) - \bar{y}]^2$ – загальна сума квадратів (total sum of squares); \bar{y} – середнє

значення; $SST = SSE + SSR$, де $SSR = \sum_{k=1}^N [\hat{y}(k) - \bar{y}]^2$ – загальна сума квадратів для регресії (sum of squares for regression) [56].

Найкращим значенням є $R^2 = 1$, коли дисперсії вимірів змінної, та цієї ж змінної, оціненої за рівнянням, збігаються. R^2 показує рівень інформативності моделі по відношенню до інформативності вибірки даних, за допомогою якої вона була оцінена.

б) Сума квадратів похибок для вибраної моделі повинна бути мінімальною, тобто,

$$\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2 \rightarrow \min_{\hat{\theta}} \quad (2.17)$$

у порівнянні з усіма іншими моделями.

7) Інформаційний критерій Акайке

$$AIC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + 2n \quad (2.18)$$

та критерій Байєса-Шварца

$$BSC = N \ln \left(\sum_{k=1}^N e^2(k) \right) + n \ln(N), \quad (2.19)$$

де $n = p + q + 1$ – число параметрів моделі, які оцінюються за допомогою статистичних даних (p – число параметрів авторегресійної частини моделі; q – число параметрів ковзного середнього; 1 з'являється тоді, коли оцінюється зміщення (або перетин), тобто a_0).

За цими критеріями вибирається та модель, для якої критерії приймають найменші значення. Введення нового регресора приводить до збільшення критерію (при цьому збільшується n), але, в той же час, зменшується сума квадратів похибок і критерій в цілому зменшується. Якщо регресор не покращує модель, то критерій збільшується. Необхідно також зазначити, що критерій Байєса-Шварца краще застосовувати для відносно великих значеннях N ($N > 100$) [56].

8) Також для визначення адекватності моделі використовується F – статистика Фішера, яка пропорційнальна відношенню:

$$F \sim \frac{R^2}{1 - R^2}, \quad (2.20)$$

а для множинної (багатофакторної) регресії вона визначається за виразом

$$F = \frac{R^2}{1 - R^2} \cdot \frac{(N - p - 1)}{p}, \quad (2.21)$$

де R^2 – коефіцієнт множинної детермінації; N – число значень ряду; p – число параметрів моделі без врахування перетину (константи).

Тому якщо $R^2 \rightarrow 1$, то $F \rightarrow \infty$. Порядок застосування F – статистики такий же, як і t – статистики. Нуль-гіпотезою є в даному випадку припущення про те, що модель неадекватна в цілому. Значення $F_{\text{крит}}$ знаходять із таблиць для F – розподілу при $(p, N - p - 1)$ степенях свободи. Якщо $F > F_{\text{крит}}$, то нуль-гіпотеза щодо неадекватності моделі в цілому відкидається на вибраному рівні значимості.

Таким чином, правильне використання методики Бокса-Дженкінса забезпечує побудову адекватної математичної моделі процесу [56].

2.3.3 Тести для аналізу даних на інтегрованість та гетероскедастичність

Для визначення наявності нестационарності застосовується тест Дікі-Фуллера. Суть тесту полягає в тому, що для визначення присутності одиничного кореня використовуються три наступні рівняння:

$$\Delta y(k) = \gamma y(k-1) + \varepsilon(k), \quad (2.22)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + \varepsilon(k), \quad (2.23)$$

$$\Delta y(k) = a_0 + \gamma y(k-1) + a_2 k + \varepsilon(k), \quad (2.24)$$

де k – дискретний час; γ – коефіцієнт у рівнянні

$$y(k) = a_0 + a_1 y(k-1) + \varepsilon(k). \quad (2.25)$$

Різниця між рівняннями (2.22) та (2.23), (2.24) полягає у присутності детермінованих членів a_0 і $a_2 k$ у рівняннях (2.23) і (2.24), відповідно. Рівняння (2.22) – це модель випадкового кроку, друге включає зсув у вигляді константи a_0 , а третє – зсув та детермінований лінійний часовий тренд. Досліджуємо параметр γ у трьох рівняннях. Якщо $\gamma = 0$, то послідовність $\{y(k)\}$ містить одиничний корінь. Застосування тесту Дікі-Фуллера передбачає оцінювання одного або більше з наведених вище трьох рівнянь за допомогою МНК або ММП з метою отримання оцінки параметра γ та стандартної похибки цієї оцінки. На основі оцінки та її стандартної похибки обчислюється t – статистика, яка порівнюється із значеннями, наведеними в таблицях Дікі-Фуллера. На основі цього порівняння приймається рішення щодо справедливості або відхилення нуль-гіпотези.

За наявності гетероскедастичності можуть бути зміщені оцінки стандартних помилок коефіцієнтів регресії, неефективні оцінки МНК коефіцієнтів регресії, неадекватна t-статистика коефіцієнтів регресії. [56]

Тому для перевірки на гетероскедастичність використовуються тести:

- тест Уайта;
- тест Голдфелда-Квандта.

Тест Уайта складається з наступних кроків:

- 1) Оцінюємо коефіцієнти основної регресії

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.26)$$

- 2) Зберігаємо залишки регресії $\hat{\varepsilon}, i = 1, \dots, n$
- 3) Оцінюємо регресію квадратів залишків квадратів залишків на всі регресори, їх квадрати, попарні добутки і константу:

$$\hat{\varepsilon}^2 = \alpha_1 + \sum_{i=2}^k \alpha_i X_i + \sum_{i=2}^k \beta_{i2} X_i^2 + \sum_{\substack{i,j=2 \\ i < j}}^k \gamma_{ij} X_i X_j + u \quad (2.27)$$

- 4) В останній оціненій регресії знаходимо коефіцієнт множинної детермінації R^2 .

- 5) Обчислюємо тестову статистику за формулою nR^2 . Якщо виконується нульова гіпотеза тестова статистика має розподіл χ^2 з $m - 1$ ступенями свободи, де m – число коефіцієнтів, які оцінювались в останній регресії.

- 6) Порівнюємо отримані значення тестової статистики з критичним значенням $\chi_\alpha^2(m - 1)$. Якщо значення тестової статистики перевищує критичне значення, тоді нульова гіпотеза про гомоскедастичність відкидається. Хоч цей

тест є універсальним, проте він не вказує на функціональну форму гетероскедастичності [57].

Тест Голдфельда-Квандта складається з наступних кроків:

- 1) Оцінюємо коефіцієнти основної регресії

$$Y = \beta_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.28)$$

- 2) Зберігаємо залишки регресії $\hat{\varepsilon}, i = 1, \dots, n$. Після аналізу графіка залишків може з'явитись припущення, що дисперсія збурень збільшується з ростом деякої змінної.

- 3) Впорядковуємо всі спостереження за модулем цієї змінної.

- 4) Ділимо всі спостереження на три групи. Зручно якщо в першій і третій групах кількість спостережень однакова.

- 5) Спостереження середньої групи нехтуємо, а по перших n_1 і останніх n_2 спостереженнях оцінюємо окремі регресії.

- 6) Використовуючи суми квадратів залишків (RSS) в оцінених регресіях, розраховуємо тестову статистику за формулою:

$$F = \frac{RSS_2 / (n_2 - k)}{RSS_1 / (n_1 - k)} \quad (2.29)$$

- 7) При виконанні нульової гіпотези тестова статистика має F-розподіл з $(n_2 - k, n_1 - k)$. Якщо значення тестової статистики перевищує критичне, то нульова гіпотеза про гомоскедастичність відкидається[56].

Висновки до розділу 2

В цьому розділі розглянуто підходи до моделювання і прогнозування лінійних та нелінійних нестационарних процесів, детально описано методику Бокса-Дженкінса, яка орієнтована на побудову моделей процесів довільної природи, в тому числі і фінансово-економічних.

Дана методика включає: загальний аналіз процесу, попередню обробку даних (нормування, корегування або виключення даних), перевірку на наявність нестационарності і нелінійності з використанням різних статистичних критеріїв, оцінювання моделей-кандидатів, вибір методу оцінювання параметрів моделей і подальша оцінка та вибір найкращої моделі за допомогою критеріїв адекватності. За наявності великої вибірки даних, яка повністю відображає поведінку і зміни в процесі, а також при коректному застосуванні цієї методики, можна побудувати адекватну математичну модель процесу.

РОЗДІЛ 3. ВИКОНАННЯ ОБЧИСЛЮВАЛЬНИХ ЕКСПЕРИМЕНТІВ

3.1. Опис програмного продукту

Лідерами серед мов програмування в сфері аналізу даних є Python та R. В даній роботі було обрано мову програмування Python як найбільш зручну для первинного аналізу даних, математичних розрахунків та візуалізації даних. Середовищем розробки було обрано Jupyter через зручність роботи.

Робота програмного продукту складається з наступних етапів:

- завантаження і обробка даних;
- аналіз даних – проведення статистичного і кореляційного аналізу, перевірка на нелінійність та нестационарність;
- побудова моделі;
- прогнозування на основі створеної моделі.

3.2. Вибір даних та моделювання і прогнозування

Для аналізу та побудови моделей і прогнозів було вибрано чотири часових рядів: акції Google (2517 значень), акції Amazon (2517 значень), CPI України (336 значень), CPI Німеччини (336 значень).

Ряд цін акцій компанії Google (Google_stocks)

На рисунку 3.1 представлено дані, що відповідають ціні закриття акції компанії Google в період з 2004 до 2019 року.

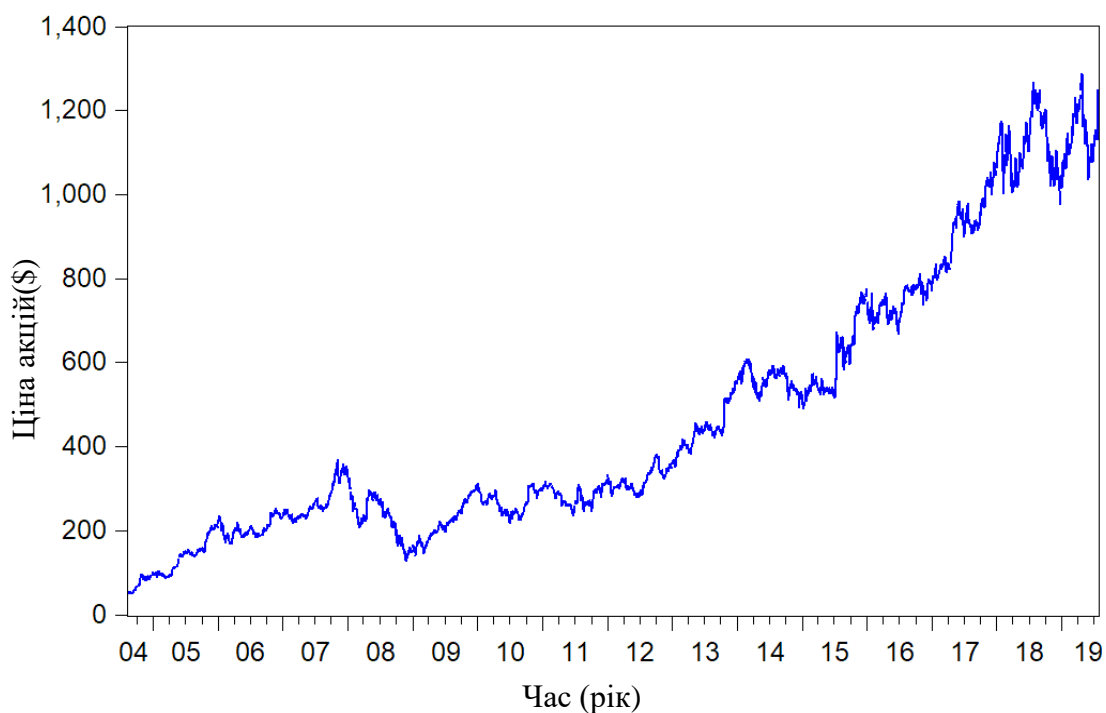


Рисунок 3.1 – Ряд Google_stocks за 2004-2019 роки.

Статистичні характеристики ряду Google_stocks:

- а) Середнє – 467,5430.
- б) Дисперсія – 317,1128.
- в) Коефіцієнт асиметрії – 0,9347.
- г) Екссес – 2,669787.
- д) Статистика Жака-Бера – 564,6122.

Опишемо вибраний процес за допомогою моделі авторегресії, оскільки авторегресійні моделі прості за своєю структурою, але досить часто вони мають високу ступінь адекватності досліджуваному процесу, прийнятну для подальшого використання.

Згідно виконаному кореляційному аналізу даних побудуємо авторегресійну модель для ряду Google_stocks, а саме авторегресію 4 порядку:

$$\text{CLOSE} = 0.226778657928 + 1.02260843726 * \text{CLOSE}(k - 1) - \\ - 0.0759979929174 * \text{CLOSE}(k - 2) + 0.0648482851956 * \text{CLOSE}(k -$$

$$-3) - 0.0112492199724 * \text{CLOSE}(k - 4) \quad (3.1)$$

Статистичні характеристики адекватності моделі:

$$R^2 = 0,91; \sum e^2 = 292,9; DW = 1,943 \quad (3.2)$$

Коефіцієнт детермінації приймає значення 0,91, сума квадратів похибок дуже велика 292,9, а статистика Дарбіна-Уотсона (1,943) наближається до найкращого значення – модель характеризується високою адекватністю.

Отримано такі характеристики однокрокового прогнозу (рис. 3.2) за побудованою моделлю:

$$\text{СКП} = 20,47; \text{САПП} = 14,0; U = 0,009 \quad (3.3)$$

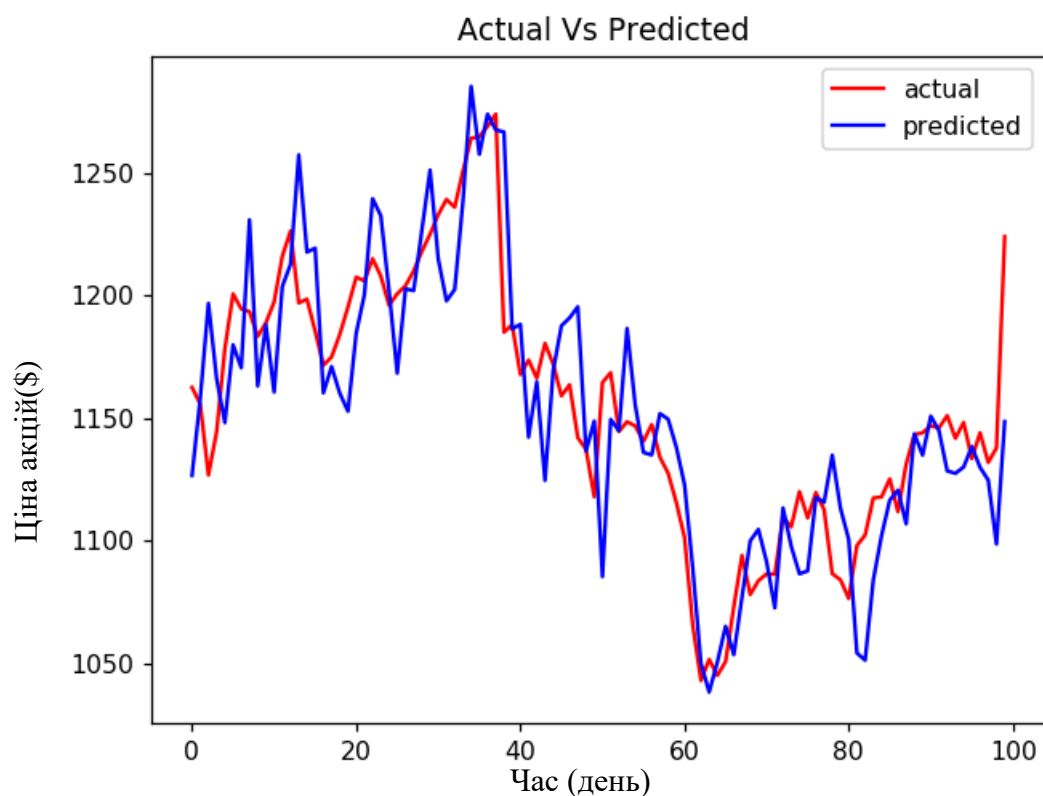


Рисунок 3.2 – Прогноз для ряду Google_stocks по моделі AP(4) (червона крива – часовий ряд, синя крива – передбачення)

За допомогою створеної моделі отримано прогноз достатньо високої якості, а коефіцієнт Тейла досягає ідеального значення (ідеальне значення = 0). Тому можна побудувати авторегресію вищого порядку, щоб покращити прогноз.

Побудуємо модель авторегресії 4-го порядку з лінійним трендом (рис. 3.3):

$$\begin{aligned} \text{CLOSE} = & -0.0413743626084 + 0.000649093926643 * T + \\ & + 1.0215032504 * \text{CLOSE}(k-1) - 0.0760121042492 * \text{CLOSE}(k-2) + \\ & + 0.0648248581303 * \text{CLOSE}(k-3) - 0.0121476168937 * \text{CLOSE}(k-4) \end{aligned} \quad (3.4)$$

Статистичні характеристики моделі:

$$R^2 = 0,97; \sum e^2 = 292,26; DW = 1,9525 \quad (3.5)$$

Отримано такі характеристики однокрокового прогнозу (рис. 3.3) за побудованою моделлю:

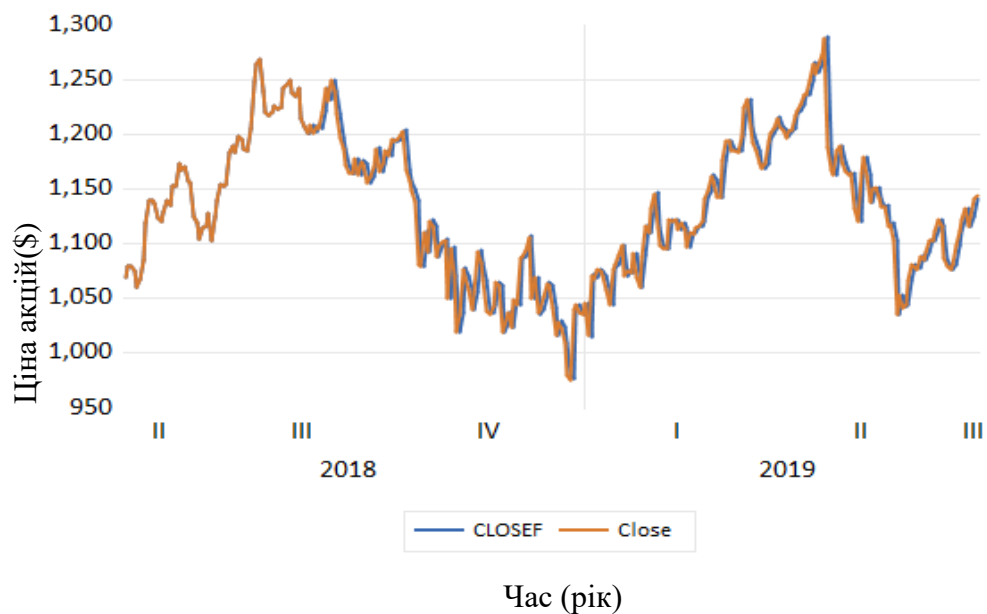


Рисунок 3.3 – Прогноз для ряду Google_stocks по моделі $AR(4) + k_2$ (червона крива – часовий ряд, синя крива – передбачення)

$$\text{СКП} = 20,34; \text{САПП} = 13,8; U = 0,008 \quad (3.6)$$

Таким чином, за допомогою побудованої моделі можна оцінити прогноз достатньо високої якості. Всі побудовані моделі для цього процесу зведені в табл. 3.1.

Наведемо загальну таблицю оцінок якості моделей та прогнозу для ряду Google _stocks (табл. 3.1).

Таблиця 3.1 – Результати моделювання і прогнозування ряду Google _stocks

Модель Процесу	Характеристики моделі			Характеристики прогнозу		
	R^2	$\sum e^2$	DW	СКП	САП	Тейла
AR(4)	0,91	292,9	1,943	20,47	14,0	0,009
AR(4) + k2	0,97	292,26	1,9525	20,34	13,8	0,008

Як видно з таблиці, найкращі результати показала модель AR(4) +k2.

Ряд цін акцій компанії Amazon (Amazon_stocks)

На рисунку 3.4 представлено дані, що відповідають ціні закриття акції компанії Amazon в період з 2011 до 2020 року.

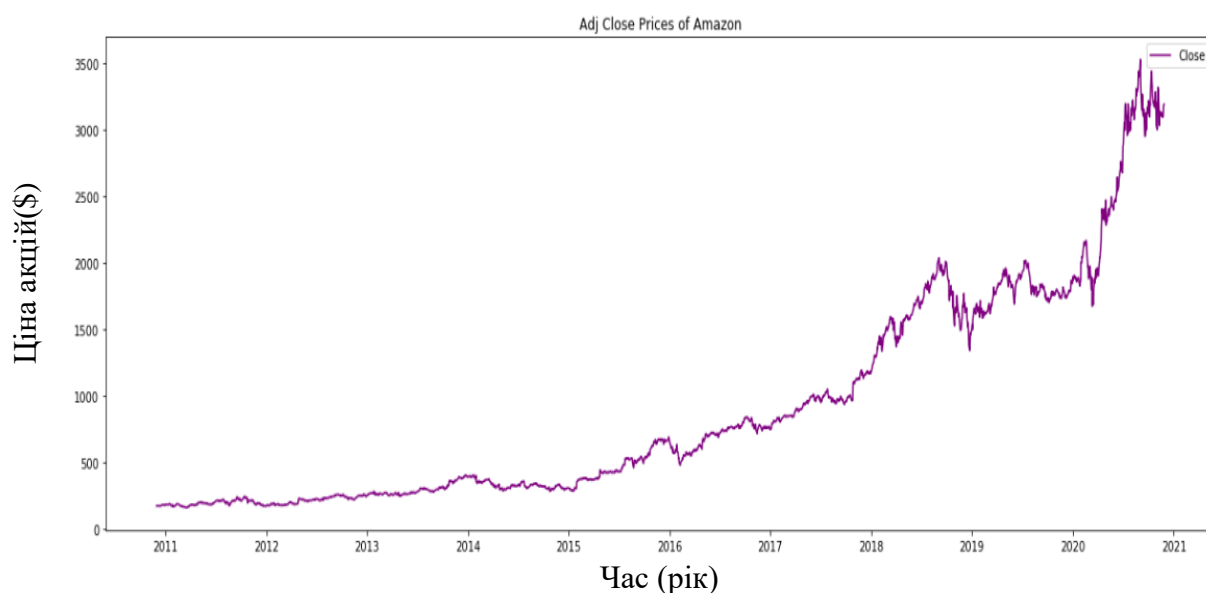


Рисунок 3.4 – Ряд Amazon_stocks

Статистичні характеристики ряду Amazon_stocks:

- а) Середнє - 903,30.
- б) Дисперсія – 796,29.
- в) Коефіцієнт асиметрії – 1,2.
- г) Екссес – 3,68.
- д) Статистика Жака-Бера – 653,19.

Побудуємо авторегресійну модель для ряду Amazon_stocks, а саме авторегресію 5 порядку:

$$\begin{aligned} \text{CLOSE} = & 0.238795535373 + 0.938904026649 * \text{CLOSE}(k - 1) + \\ & + 0.0764219828864 * \text{CLOSE}(k - 2) - 0.111698018671 * \text{CLOSE}(k - 3) + \\ & + 0.0986171998667 * \text{CLOSE}(k - 4) - 0.000992113508549 * \text{CLOSE}(k - 5) \end{aligned} \quad (3.7)$$

Статистичні характеристики моделі:

$$R^2 = 0,54; \sum e^2 = 1584,36; DW = 1,99 \quad (3.8)$$

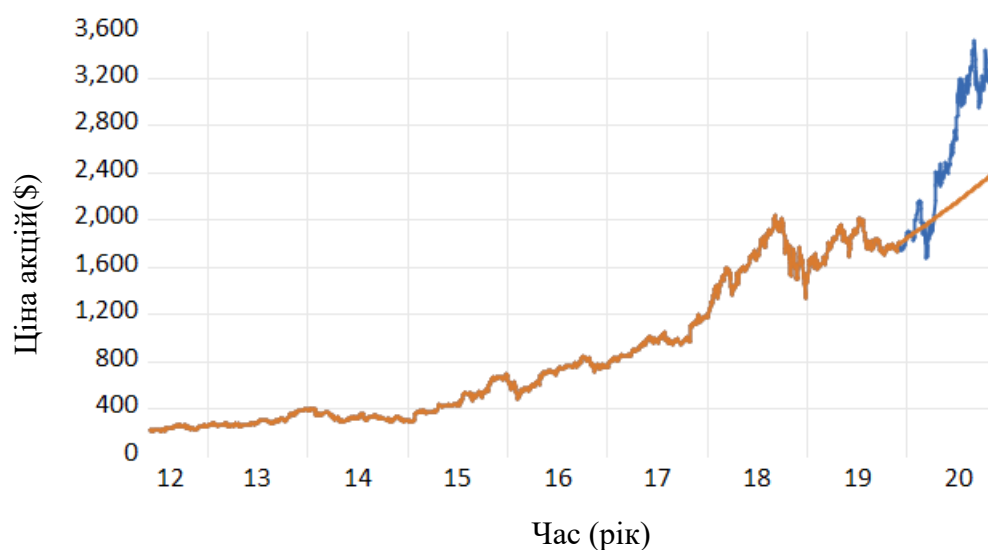


Рисунок 3.5 – Прогноз ряду Amazon_stocks по моделі AP(5) (синя крива – часовий ряд, червона крива – передбачення)

Отримано такі характеристики однокрокового прогнозу (рис. 3.5) за побудованою моделлю:

$$\text{СКП} = 51,3; \text{САПП} = 42,9; U = 0,002 \quad (3.9)$$

Побудуємо модель авторегресії 5-го порядку з квадратичним трендом:

$$\begin{aligned} \text{CLOSE} = & 0.124311758503 + 2.86175368249\text{e-}06 * T1 + \\ & + 0.935788508446 * \text{CLOSE}(k - 1) + 0.0758735638727 * \text{CLOSE}(k - 2) - \\ & - 0.111651552721 * \text{CLOSE}(k - 3) + 0.0981638832469 * \text{CLOSE}(k - 4) - \\ & - 0.00350345485163 * \text{CLOSE}(k - 5) \end{aligned} \quad (3.10)$$

Статистичні характеристики моделі:

$$R^2 = 0,99; \sum e^2 = 1579,4; DW = 2,0 \quad (3.11)$$

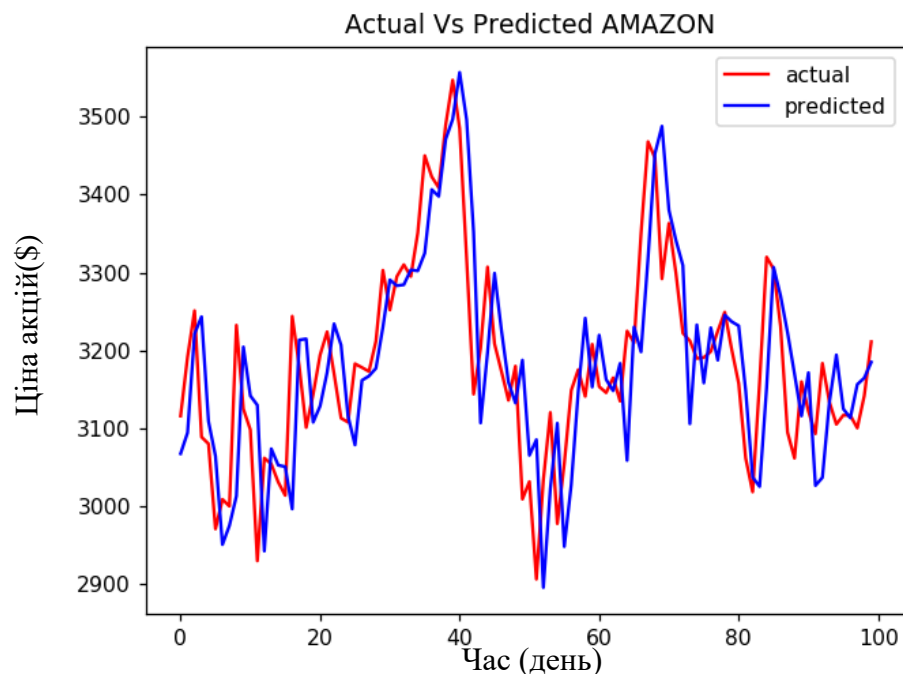


Рисунок 3.6 – Прогноз ряду Amazon_stocks по моделі AP(5) з квадратичним трендом (червона крива – часовий ряд, синя крива – передбачення).

Отримано такі характеристики однокрокового прогнозу (рис. 3.6) за

побудованою моделлю:

$$\text{СКП} = 49,3; \text{САП} = 35,9; U = 0,002 \quad (3.12)$$

Наведемо загальну таблицю оцінок якості моделей та прогнозу для ряду Amazon_stocks (табл. 3.2).

Таблиця 3.2 – Результати моделювання і прогнозування ряду Amazon_stocks

Модель процесу	Характеристики моделі			Характеристики прогнозу		
	R^2	$\sum e^2$	DW	СКП	САП	Тейла
AP(5)	0,54	1584,36	1,99	51,3	42,9	0,002
AP(5) +k2	0,99	1579,4	2,0	49,3	35,9	0,002

Найкращі результати показала модель AP(5) +k2.

Ряд індексів споживчих цін України (CPI_Ukr)

На рисунку 3.7 представлено дані, що відповідають індексу споживчих цін в Україні в період з 1994 до 2020 року. Індекс споживчих цін (ІСЦ) – це показник, який характеризує зміни у часі загального рівня цін на товари та послуги, які купує населення для власного невиробничого споживання. ІСЦ використовують у якості показника загального рівня інфляції в економіці, тому його називають «індексом інфляції».

Статистичні характеристики ряду CPI_Ukr:

- а) Середнє – 103,7423.
- б) Дисперсія – 11,29229.
- в) Коефіцієнт асиметрії – 5,279723.
- г) Екссес – 33,28448.
- д) Статистика Жака-Бера – 14401,13.

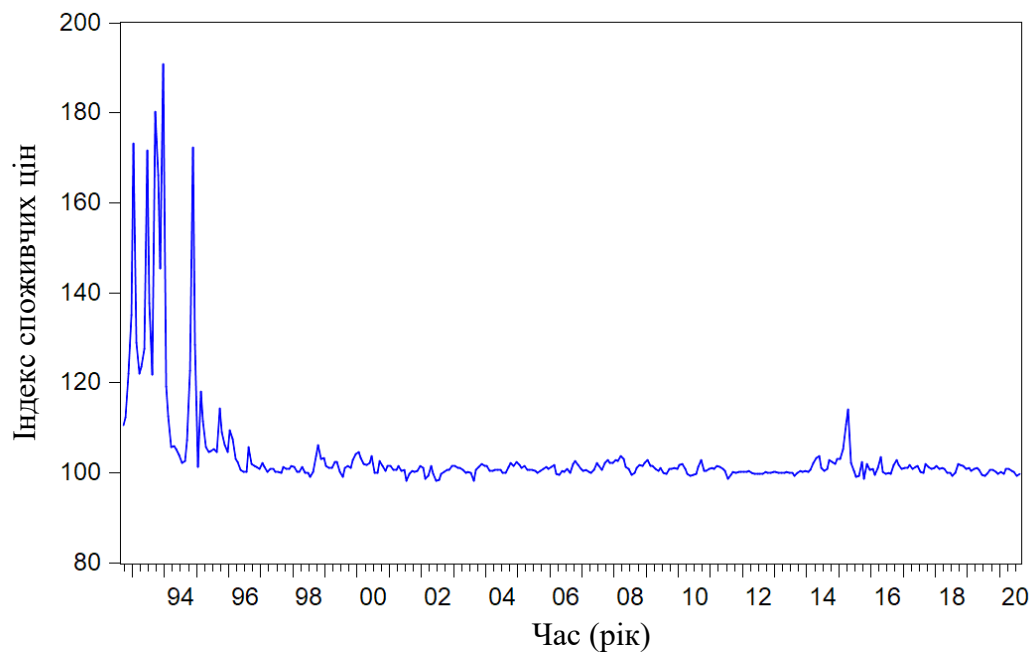


Рисунок 3.7 – Ряд CPI_Ukr

Побудуємо авторегресійну модель для ряду CPI_Ukr, авторегресію 5 порядку:

$$\begin{aligned} \text{DATA} = & 0.522292076214 * \text{DATA}(k-1) - 0.0593932508933 * \text{DATA}(k-2) + \\ & + 0.363663289085 * \text{DATA}(k-3) - 0.21927957845 * \text{DATA}(k-4) + \\ & + 0.201738323367 * \text{DATA}(k-5) + 19.4964985809 \end{aligned} \quad (3.13)$$

Статистичні характеристики моделі:

$$R^2 = 0,595951; \sum e^2 = 14707,42; DW = 1,863146. \quad (3.14)$$

Коефіцієнт детермінації приймає значення 0,595951, сума квадратів похибок становить 14707,42, а статистика Дарбіна-Уотсона (1,863146) наближається до найкращого значення.

Отримано такі характеристики однокрокового прогнозу (рис. 3.8) за побудованою моделлю:

$$\text{СКП} = 0,70; \text{САПП} = 0,57; U = 0,0035 \quad (3.15)$$

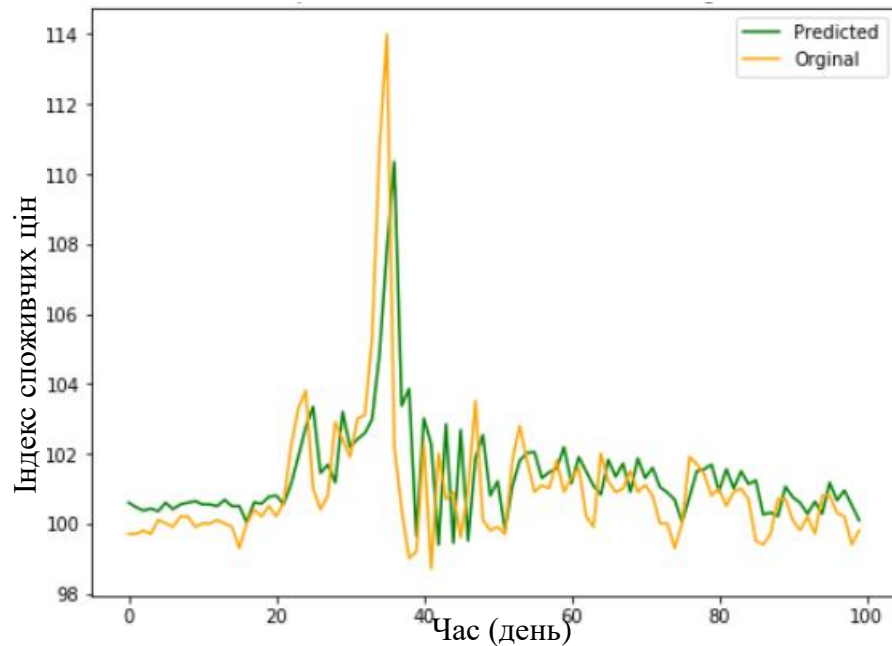


Рисунок 3.8 – Прогноз ряду CPI_Ukr по моделі AP(5) (жовта крива – часовий ряд, зелена крива – передбачення)

За допомогою створеної моделі отримано прогноз високої якості, а коефіцієнт Тейла (ідеальне значення = 0) підтверджує можливість покращення прогнозу. Тому необхідно побудувати авторегресію вищого порядку.

Авторегресія 14-го порядку:

$$\begin{aligned} \text{DATA} = & 29.25636991 + 0.420828143915 * \text{DATA}(k - 1) + \\ & + 0.0963893030193 * \text{DATA}(k - 2) + 0.208411066064 * \text{DATA}(k - 3) - \\ & - 0.168044662463 * \text{DATA}(k - 4) - 0.0631331891297 * \text{DATA}(k - 5) - \\ & - 0.0137722271709 * \text{DATA}(k - 10) + 0.413445783205 * \text{DATA}(k - 11) - \\ & - 0.155298036964 * \text{DATA}(k - 12) - 0.0287727989217 * \text{DATA}(k - 14) \end{aligned} \quad (3.16)$$

$$R^2 = 0,676652; \sum e^2 = 5809,241; DW = 2,045 \quad (3.17)$$

Коефіцієнт детермінації приймає значення 0,676652, сума квадратів похибок становить 5809,241, а статистика Дарбіна-Уотсона досягає найкращого значення.

Отримано такі характеристики однокрокового прогнозу (рис. 3.9) за побудованою моделлю:

$$\text{СКП} = 0,67; \text{САПП} = 0,53; U = 0,0033 \quad (3.18)$$

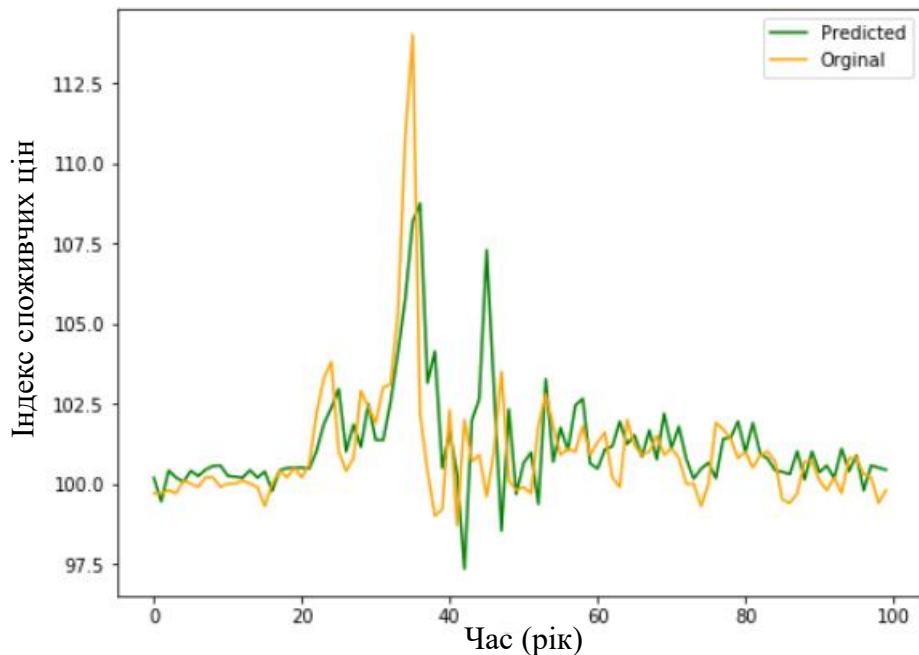


Рисунок 3.9 – Прогноз ряду CPI_Ukr по моделі AP(14) (жовта крива – часовий ряд, зелена крива – передбачення).

Далі побудуємо іншу авторегресійну модель для ряду CPI_Ukr: авторегресія 14 порядку з ковзним середнім порядку 8:

$$\begin{aligned} D1 = & 51.9661526904 + 0.375046253364 * \text{DATA}(k - 1) - \\ & - 0.320936448877 * \text{DATA}(k - 2) + 0.0807534418363 * \text{DATA}(k - 3) - \\ & - 0.221742961287 * \text{DATA}(k - 4) + 0.0785096489232 * \text{DATA}(k - 5) + \\ & + 0.0951205112603 * \text{DATA}(k - 10) + 0.248561938921 * \text{DATA}(k - 11) - \\ & - 0.0630473631917 * \text{DATA}(k - 12) + 0.211871289848 * \text{DATA}(k - 4) - \\ & - 0.0642107064596 * \text{MV}(k - 1) - 0.95074479995 * \text{MV}(k - 2) + \\ & + 1.41093050825 * \text{MV}(k - 3) - 1.39126715147 * \text{MV}(k - 4) + \\ & + 1.99572826254 * \text{MV}(k - 5) - 1.88370181438 * \text{MV}(k - 6) + \end{aligned}$$

$$+0.6074382547 * MV(k-7) - 0.19165567418 * MV(k-8) \quad (3.19)$$

$$R^2 = 0,88; \sum e^2 = 758,0647; DW = 2,074 \quad (3.20)$$

Отримано такі характеристики однокрокового прогнозу (рис. 3.10) за побудованою моделлю:

$$СКП = 0,66; САПП = 0,57; U = 0,0033 \quad (3.21)$$

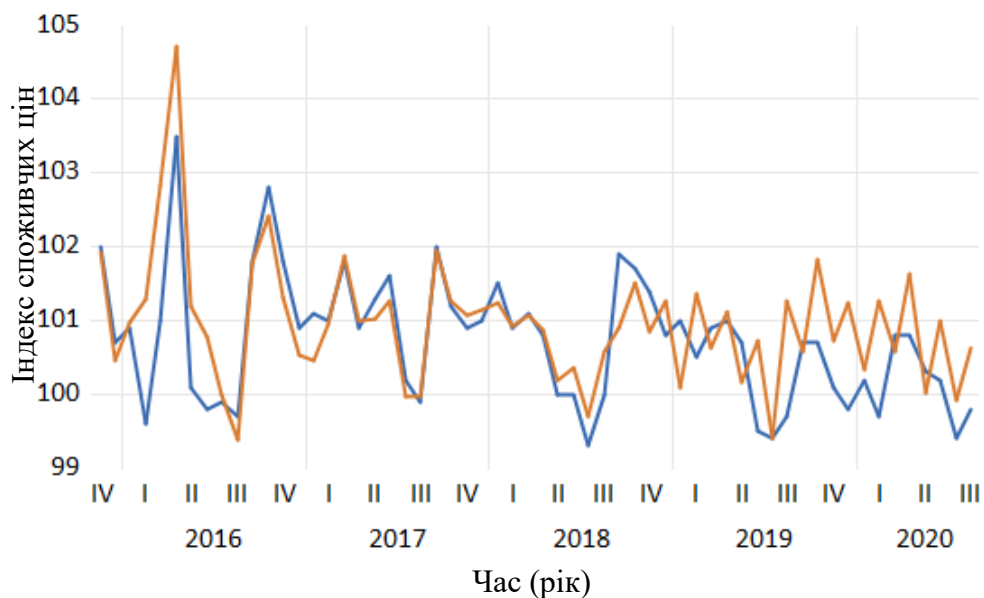


Рисунок 3.10 – Прогноз ряду CPI_Ukr по моделі АРКС (14, 8) часовий ряд (синя крива – часовий ряд, червона крива – передбачення)

Таким чином, за допомогою побудованої моделі можна оцінити прогноз достатньо високої якості. Всі побудовані моделі для цього процесу зведені в табл. 3.3.

Наведемо загальну таблицю оцінок якості моделей та прогнозу для ряду CPI_Ukr (табл. 3.3).

Таблиця 3.3 – Результати моделювання і прогнозування ряду CPI_Ukr

Модель процесу	Характеристики моделі			Характеристики прогнозу		
	R^2	$\sum e^2$	DW	СКП	САП	Тейла
AP(5)	0,596	14707,42	1,86	0,70	0,57	0,0035
AP(14)	0,677	5809,24	2,045	0,67	0,53	0,0033
АРКС(14, 8)	0,88	758,05	2,074	0,66	0,57	0,0033

Найкращі результати показала модель АРКС(14,8).

Ряд індексів споживчих цін Німеччини (CPI_Germany)

На рисунку 3.11 представлено дані, що відповідають індексу споживчих цін у Німеччині в період з 1994 до 2020 року.

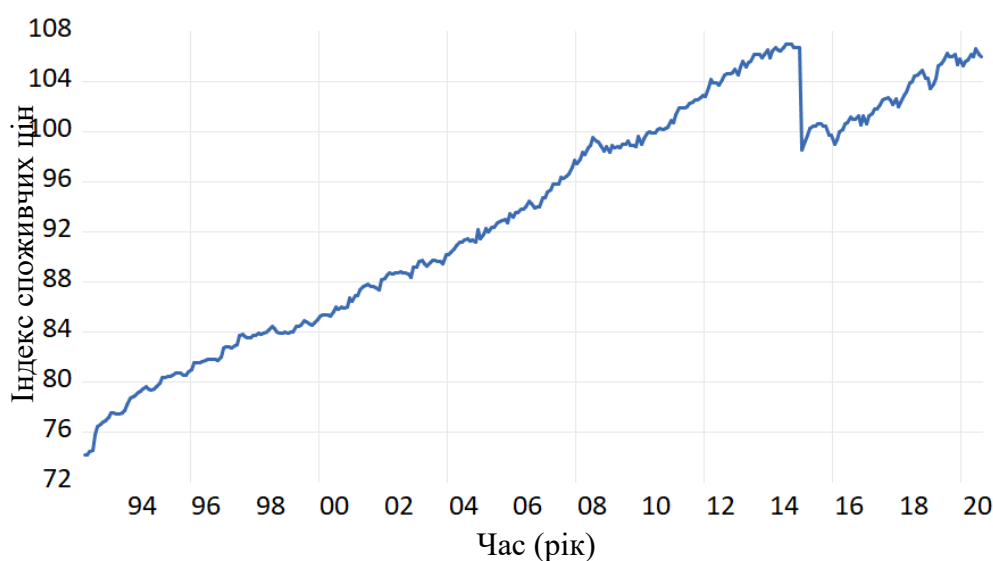


Рисунок 3.11 – Ряд CPI_Germany в період з 1994 до 2020 року

Статистичні характеристики ряду CPI_Germany:

- а) Середнє – 93,48.
- б) Дисперсія – 9,35.
- в) Коефіцієнт асиметрії – -0,23.
- г) Екссес – 1,73.

д) Статистика Жака-Бера – 25,63.

Побудуємо авторегресійну модель для ряду CPI_Germany: авторегресія 2 порядку:

$$\begin{aligned} \text{DATA} = & 0.649870458232 + 0.890397636745 * \text{DATA}(k - 1) + \\ & + 0.103779966676 * \text{DATA}(k - 2) \end{aligned} \quad (3.22)$$

$$R^2 = 0,99; \sum e^2 = 100,56; DW = 1,995 \quad (3.23)$$

Отримано такі характеристики однокрокового прогнозу (рис. 3.12) за побудованою моделлю:

$$\text{СКП} = 0,45; \text{САПП} = 0,37; U = 0,002 \quad (3.24)$$

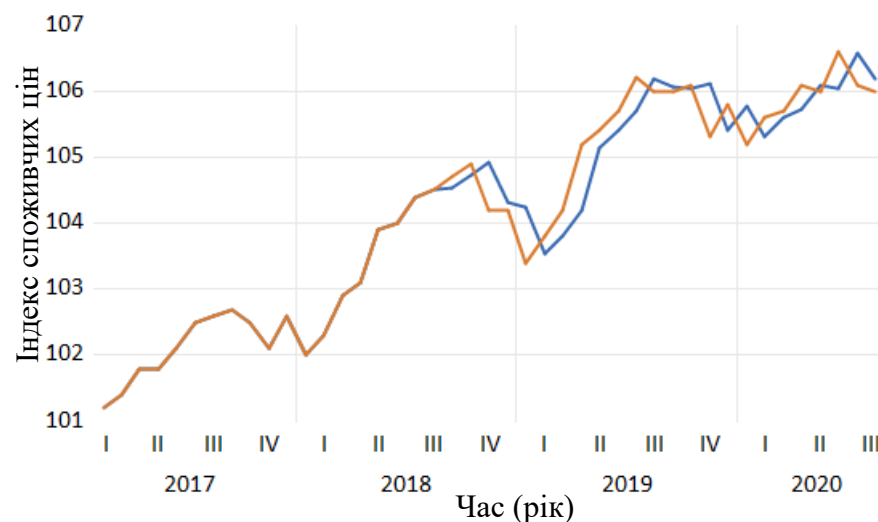


Рисунок 3.12 – Прогноз ряду CPI_Germany по моделі AP(2) (червона крива – часовий ряд; синя крива – передбачення)

Таким чином, за допомогою побудованої моделі можна оцінити прогноз достатньо високої якості.

Наведемо загальну таблицю оцінок якості моделей та прогнозу для ряду

CPI_ Germany (табл. 3.4).

Таблиця 3.4 – Результати моделювання і прогнозування ряду CPI_ Germany

Модель Процесу	Характеристики моделі			Характеристики прогнозу		
	R^2	$\sum e^2$	DW	СКП	САПП	Тейла
AR(2)	0,99	100,56	1,995	0,45	0,37	0,002

На основі результатів попередніх досліджень побудуємо моделі обраних процесів за допомогою програмного продукту GMDH Shell який використовує метод групового аргументу для побудови моделей.

Проаналізуємо та спрогнозуємо ряд Google_stocks за допомогою МГУА. Отримаємо модель:

$$N2 = 0.811887 + \text{open} * 0.989706 + \text{open} * \text{"open, cubert"} * 0.00101761 \quad (3.25)$$

Коефіцієнт детермінації приймає значення $R^2 = 0,93$.

Зобразимо графік прогнозу для обраного ряду (рис. 3.13).

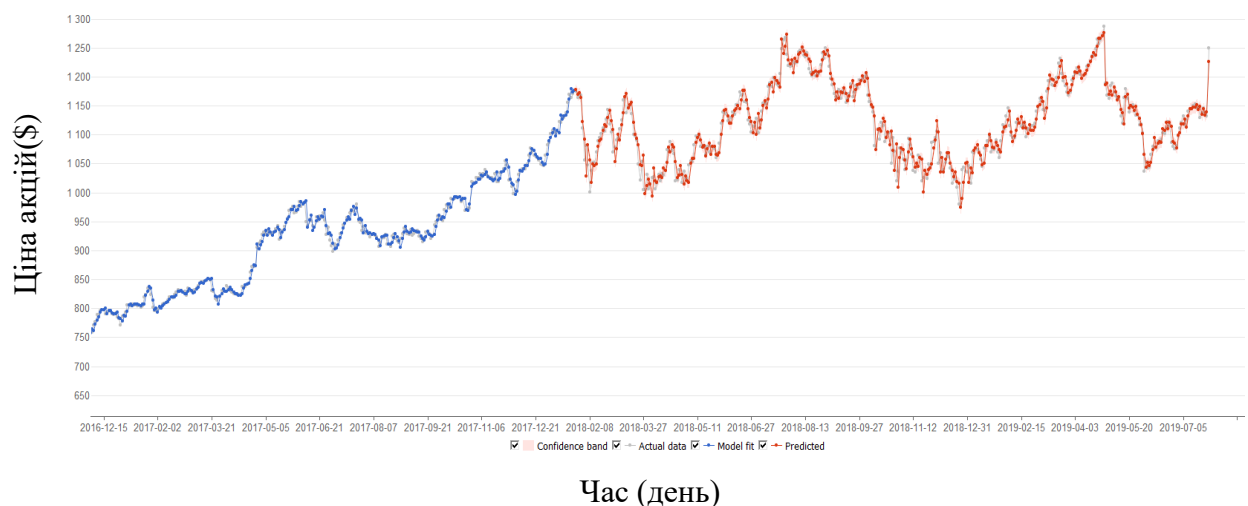


Рисунок 3.13 – МГУА Google_stocks (сіра крива – часовий ряд; червона крива – передбачення; синя крива - моделювання)

Характеристика прогнозу даного ряду приймає такі значення:

$$\text{СКП} = 1,49 \% ; \text{САПП} = 1,08\% \quad (3.26)$$

Показники прогнозу дуже гарні, середня абсолютна похибка в процентах (САПП) склала всього 1,08%. Можна зробити висновок, що модель цілком придатна для використання.

Наведемо модель яка побудована для ряду AMAZON_stocks:

$$\begin{aligned} N2 = & -30.2442 + \text{open} * \text{"open, cubert"} * 0.0679473 - \\ & - \text{open}^2 * 0.000101966 + \text{"open, cubert"}^2 * 4.53397 \end{aligned} \quad (3.27)$$

Коефіцієнт детермінації приймає значення $R^2 = 0,98$.

Зобразимо графік прогнозу для обраного ряду (рис. 3.14).

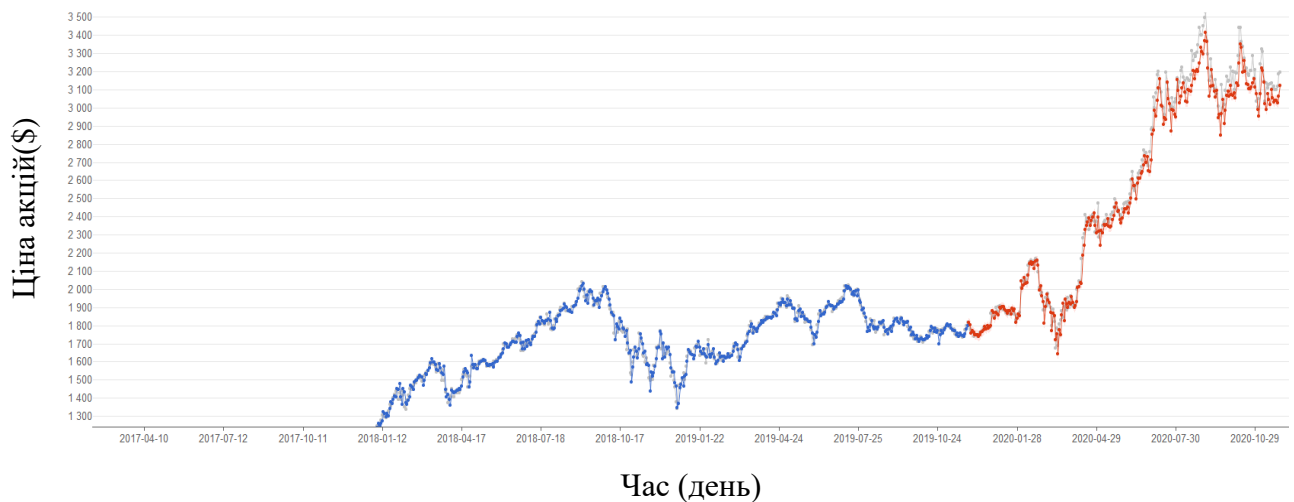


Рисунок 3.14 – МГУА AMAZON_stocks (сіра крива – часовий ряд; червона крива – передбачення; синя крива - моделювання)

Характеристика прогнозу даного ряду приймає наступні значення:

$$\text{СКП} = 2,42 \% ; \text{САПП} = 1,91\%. \quad (3.28)$$

Показники прогнозу дуже гарні, середня абсолютна похибка в процентах (САПП) склала всього 1,91%. Можна зробити висновок, що модель цілком придатна для використання.

Наведемо модель яка побудована для ряду CPI_Ukr:

$$N8 = 72.5424 + \text{"Data[t-3], cubert"} * 8.26123 + \text{"Data[t-3], cubert"} * \text{cycle} * 0.198031 - \text{cycle} * 0.394336 + \text{cycle}^2 * 0.00633946 \quad (3.29)$$

Коефіцієнт детермінації приймає значення $R^2 = 0,67$

Зобразимо графік прогнозу для обраного ряду (рис. 3.15).

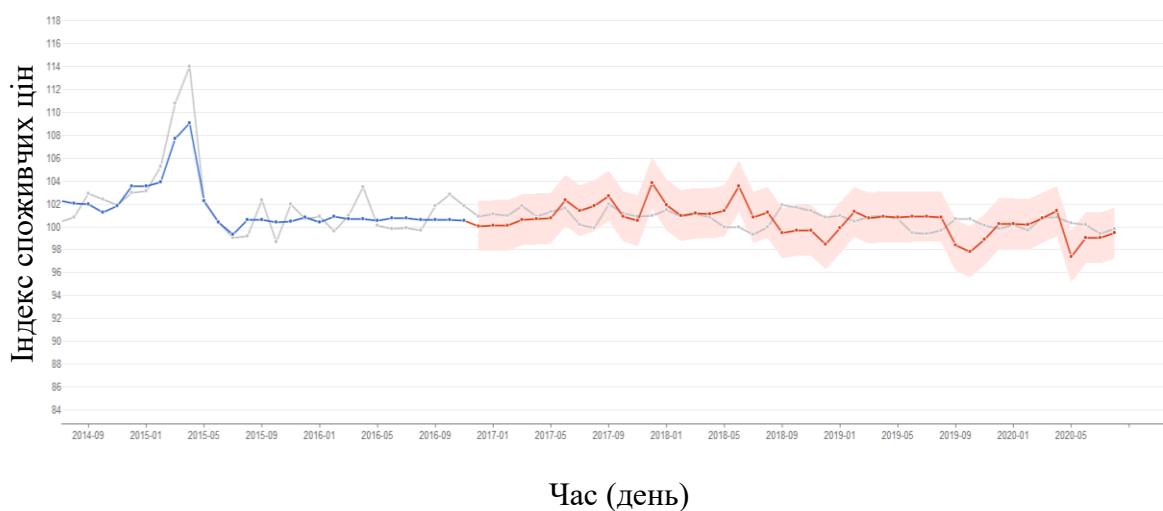


Рисунок 3.15 – МГУА CPI_Ukr (сіра крива – часовий ряд; червона крива – передбачення; синя крива - моделювання)

Характеристика прогнозу даного ряду приймає наступні значення:

$$\text{СКП} = 1,4 \% ; \text{САПП} = 1,07\% . \quad (3.30)$$

Наведемо модель яка побудована для ряду CPI_Germany:

$$Y1 = -168.411 + \text{"Data[t-10], cubert"} * 57.9576 \quad (3.31)$$

Коефіцієнт детермінації приймає значення $R^2 = 0,87$

Зобразимо графік прогнозу для обраного ряду (рис. 3.16).

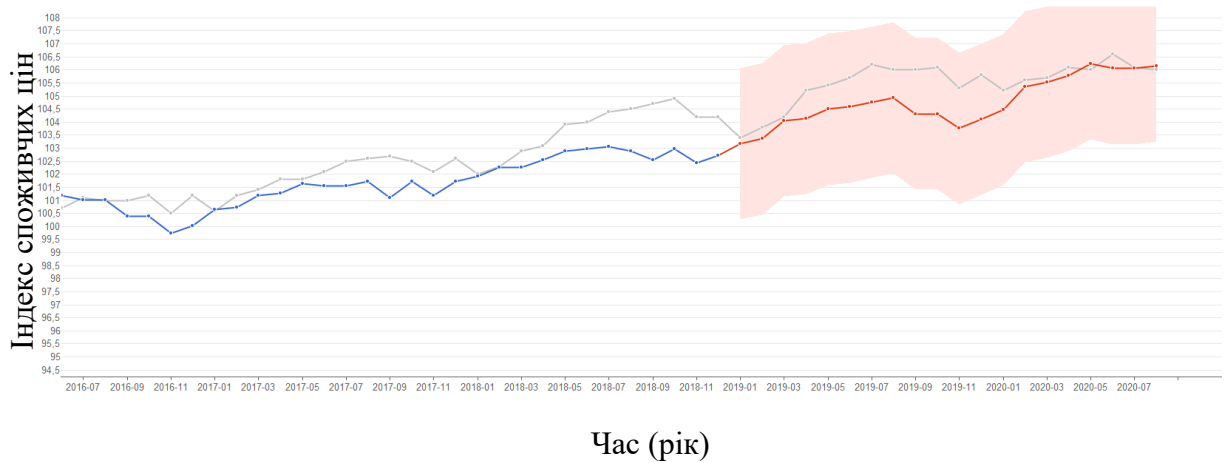


Рисунок 3.16 – МГУА CPI_Germany (сіра крива – часовий ряд; червона крива – передбачення; синя крива - моделювання)

Характеристика прогнозу даного ряду приймає наступні значення:

$$\text{СКП} = 1,5 \% ; \text{САПП} = 0,85 \% . \quad (3.32)$$

Наведемо загальну таблицю оцінок якості моделей та прогнозу для всіх рядів (табл. 3.5).

Таблиця 3.5 – зведена таблиця оцінок обраних рядів

Ряд	Тип моделі	Адекватність моделі			Характеристика прогнозу	
		R^2	$\sum e^2$	DW	СКП	Тейла
Google stocks	AP(4)	0,91	292,2	0,943	20,47	0,009
	AP (4)+k2	0,97	292,26	1,9525	20,34	0,008
	МГУА	0,93			1,49	
Amazon stocks	AP(5)	0,54	1584,36	1,99	51,3	0,002
	AP(5)+k2	0,99	1579,4	2,0	49,3	0,002
	МГУА	0,98			2,42	

Продовження таблиці 3.5

CPI Ukraine	AP(5)	0,596	14707,4	1,86	0,70	0,0035
	AP(14)	0,677	5809,24	2,045	0,67	0,0033
	АРКС(14,8)	0,88	758,05	2,074	0,57	0,0033
	МГУА	0,67			1,4	
CPI Germany	AP(5)	0,99	100,56	1,995	0,45	0,002
	МГУА	0,87			1,5	

Висновки до розділу 3

У даному розділі здійснено моделювання і прогнозування розвитку вибраних процесів на основі декількох моделей: авторегресійних моделей (АР), авторегресійних моделей з ковзним середнім (АРКС) та методу групового врахування аргументів. Після виконання порівняльного аналізу було обрано найкращу модель для кожного ряду.

Створено програмний продукт (мовою програмування Python), за допомогою якого можна реалізувати тест Дікі-Фулера, видаляти тренди за допомогою різниць (перших або других), будувати графіки функцій автокореляції та часткової автокореляції, створювати ARMA/ARIMA моделі та здійснювати прогнозування.

Було показано, що варіюванням порядку регресії та порядку ковзного середнього можна досягти мінімального значення середньої квадратичної похибки. Результати моделювання і прогнозування за допомогою різних методів свідчать про те, що не існує єдиного універсального методу. Для кожного часового ряду потрібно використовувати різні підходи та враховувати параметри і критерії за допомогою яких можна побудувати найбільш адекватну модель та виконати якісний прогноз.

РОЗДІЛ 4 РОЗРОБКА СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

В межах підпункту було проаналізовано:

- зміст ідеї (що пропонується);
- можливі напрямки застосування;
- основні вигоди, що може отримати користувач товару (за кожним напрямком застосування);
- чим відрізняється від існуючих аналогів та замінників;

Перші три пункти подані у вигляді таблиці (табл. 4.1) і дають цілісне уявлення про зміст ідеї та можливі базові потенційні ринки, в межах яких потрібно шукати групи потенційних клієнтів.

Таблиця 4.1 - Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Розроблену систему моделювання процесів та побудови прогнозів можна використовувати для аналізу та прогнозування нелінійних нестаціонарних процесів в економіці та фінансах.	В державних установах з метою побудови моделей та прогнозів, виявлення можливих ризиків.	Більшість процесів в економіці та фінансах є складними, оскільки вони нелінійні та нестаціонарні, тому необхідно створювати нові моделі для підвищення якості прогнозів.
	На підприємствах з метою прийняття рішень на основі результатів роботи програми.	За допомогою якісного прогнозу можна планувати подальший розвиток підприємства.

Аналіз потенційних техніко-економічних переваг ідеї (чим відрізняється від існуючих аналогів та замінників) порівняно із пропозиціями конкурентів передбачає:

- визначення переліку техніко-економічних властивостей та характеристик ідеї;
- визначення попереднього кола конкурентів (проектів-конкурентів) або товарів-замінників чи товарів-аналогів, що вже існують на ринку, та проводиться збір інформації щодо значень техніко-економічних показників для ідеї власного проекту та проектів-конкурентів відповідно до визначеного вище переліку;
- проводиться порівняльний аналіз показників: для власної ідеї визначаються показники, що мають а) гірші значення (W, слабкі); б) аналогічні (N, нейтральні) значення; в) кращі значення (S, сильні) (табл. 4.2).

Таблиця 4.2 - Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/ п	Техніко- економічні характеристик и ідеї	(потенційні) товари/концепції конкурентів		W (слабка сторона)	N (нейтральн а сторона)	S (сильна сторона)
		Мій проект	Конкурент 1			
1.	Можливість використання на різних операційних системах	Windows, Mac, Linux	Тільки Windows			+
2	Зручність у використанні	+	+		+	

Продовження таблиці 4.2

3.	Наявність різних функцій	Лише побудова моделі та прогнозування	+	+		
----	--------------------------	---------------------------------------	---	---	--	--

Визначений перелік слабких, сильних та нейтральних характеристик та властивостей ідеї потенційного товару є підґрунтям для формування його конкурентоспроможності.

4.2 Технологічний аудит ідеї стартап-проекту

В межах даного підрозділу було проведено аудит технології, за допомогою якої можна реалізувати ідею проекту (технології створення товару). Визначення технологічної здійсненності ідеї проекту передбачає аналіз таких складових (таблиця 4.3):

- за якою технологією буде виготовлено товар згідно ідеї проекту?
- чи існують такі технології, чи їх потрібно розробити/додати?
- чи доступні такі технології авторам проекту?

Таблиця 4.3 - Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1.	Створення програмного продукту для моделювання і прогнозування процесу	Алгоритми аналізу на нелінійність та нестационарність	Наявна	Доступна
		Python	Наявна	Доступна
		Eviews	Наявна	Доступна

Обрана технологія реалізації ідеї проекту: Алгоритми + Python + Eviews.

За результатами аналізу таблиці зроблено висновок щодо можливості технологічної реалізації проекту. Технологічним шляхом реалізації проекту було обрано PyCharm і Jupyter Notebook через їх доступність, безкоштовність та можливість використовувати бібліотеки та модулі для моделювання, прогнозування та візуалізації.

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Визначення ринкових можливостей, які можна використати під час ринкового впровадження проекту, та ринкових загроз, які можуть перешкодити реалізації проекту, дозволяє спланувати напрями розвитку проекту із урахуванням стану ринкового середовища, потреб потенційних клієнтів та пропозицій проектів-конкурентів.

Спочатку було проведено аналіз попиту: наявність попиту, обсяг, динаміка розвитку ринку (таблиця 4.4).

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	3
2	Загальний обсяг продаж, грн/ум.од	150000
3	Динаміка ринку (якісна оцінка)	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Немає
5	Специфічні вимоги до стандартизації та сертифікації	Немає

Продовження таблиці 4.4

6	Середня норма рентабельності в галузі (або по ринку), %	30
---	---	----

Середню норму рентабельності в галузі було порівняно із банківським відсотком на вкладення. Останній є меншим, тому є сенс вкладати гроші саме у цей проект.

За результатами аналізу таблиці 4.4 було зроблено висновок, що ринок є привабливим для входження.

Надалі були визначені потенційні групи клієнтів, їх характеристики та сформовано орієнтовний перелік вимог до товару для кожної групи (табл. 4.5).

Таблиця 4.5 - Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1.	Програмний продукт для моделювання нелінійних нестационарних процесів та їхнього прогнозування.	Державні установи, компанії, банки, біржі	Різні підходи до прогнозування та моделювання процесів.	Зручний інтерфейс, швидкість обрахунків, висока точність, можливість роботи з великими об'ємами даних різних форматів.

Після визначення потенційних груп клієнтів було проведено аналіз ринкового середовища: складено таблиці факторів, що сприяють ринковому впровадженню проекту, та факторів, що йому перешкоджають (табл. 4.6, 4.7).

Таблиця 4.6 - Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Нестача технічних ресурсів	Користувачі можуть мати обмежені технічні ресурси, яких недостатньо для повноцінної роботи	Винесення модуля обчислень на сервери компаній-партнерів
2	Зміна потреб користувачів	Користувачам необхідний програмний продукт з більшими функціональними можливостями	Постійне оновлення програмного продукту і додавання нового функціоналу
3	Конкуренція	Поява нових програмних продуктів, з кращим інтерфейсом і функціоналом	Розширення цільової аудиторії, розширення функціоналу програмного продукту

Таблиця 4.7 - Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Коригування прогнозу	Можливість коригувати прогноз за рахунок оновлення бази даних	Розробка модулів інтеграції з обліковими системами

Продовження таблиці 4.7

2	Поява нових методів моделювання	Поява методів, які будуть швидше і більш точно моделювати процеси	Покращення програмного продукту, розширення функціоналу
---	---------------------------------	---	---

Надалі було проведено аналіз пропозиції: визначили загальні риси конкуренції на ринку (таблиця 4.8).

Таблиця 4.8 - Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції - олігополія	Присутня невелика кількість фірм. Більшість ринку контролюють фірми-гіганти	Підтримка якості продукту та постійні оновлення, вдосконалення.
2. За рівнем конкурентної боротьби – міжнародний	Компанії-конкуренти з інших країн	Створити програмний продукт таким чином, щоб можна було переробити для використання у галузях інших країн.
3. За галузевою ознакою - міжгалузева	Продукт може використовуватись для різних галузей	Постійне вдосконалення продукту, що не має прив'язки до сфери

Продовження таблиці 4.8

4. Конкуренція за видами товарів: - товарно-видова	Конкуренція між видами програмного продукту, їх особливостями.	Створити програмний продукт, враховуючи недоліки конкурентів
5. За характером конкурентних переваг - нецінова	Вдосконалення технології розробки програмного продукту для зменшення собівартості	Удосконалення моделі. Використання більш дешевих технологій для розробки, ніж використовують конкуренти, але тільки якщо ці технології відповідають необхідним вимогам якості.
6. За інтенсивністю - не марочна	Торгова марка майже немає впливу	Реклама, участь у конференціях, семінарах.

Було проведено аналіз конкуренції у галузі за моделлю М. Портера (табл. 4.9).

За результатами аналізу табл. 4.9 зроблено висновок про можливість роботи на ринку з огляду на конкурентну ситуацію. Також зроблено висновок щодо характеристик, які повинен мати проект, щоб бути конкурентноспроможним на ринку.

Цей висновок був врахований при формулюванні переліку факторів конкурентноспроможності у наступному пункті. На основі аналізу конкуренції, проведеного в табл. 4.9, а також із урахуванням характеристик ідеї проекту (табл. 4.2), вимог споживачів до товару (табл. 4.5) та факторів маркетингового середовища (таблиці 4.6, 4.7) визначається та обґрунтовується перелік факторів конкурентноспроможності. Аналіз оформлено у табл. 4.10.

Таблиця 4.9 - Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари-замінники
	Навести перелік прямих конкурентів	Визначити бар'єри входження в ринок	Визначити фактори сили постачальників	Визначити фактори сили споживачів	Фактори загроз з боку замінників
	SAS Matlab STATISTIC A	Наявність вже існуючих рішень	-	Контроль якості продукту	Наявність більш широкого функціоналу, більш зручного інтерфейсу та авторитет (перевірена якість)
Висновки:	Досить інтенсивна конкуренція на ринку з існуючими компаніями	Є можливості виходу на ринок, але є і конкуренти.	-	Клієнти диктують умови роботи на ринку: зручний інтерфейс, надійний, швидкий.	Необхідно випускати програмний продукт не гірший, ніж у конкуренті в та розширяти функціонал

Таблиця 4.10 - Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування (наведення чинників, що роблять фактор для порівняння конкурентних проєктів значущим)
1	Ціна	Більш доступна ціна збільшує кількість потенційних клієнтів
2	Кросплатформність програмного продукту	Можливість використання програмного продукту на будь-якій платформі.
3	Орієнтованість на кінцевого споживача	Продукт орієнтований на взаємодію з клієнтом

За визначеними факторами конкурентоспроможності (табл. 4.10) проведено аналіз сильних та слабких сторін стартап-проєкту (табл. 4.11).

Таблиця 4.11 - Порівняльний аналіз сильних та слабких сторін проєкту

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з ... (назва підприємства)						
			-3	-2	-1	0	1	2	3
1	Ціна	15					*		
2	Кросплатформність програмного продукту	20			*				
3	Орієнтованість на кінцевого споживача	7					*		

Фінальним етапом ринкового аналізу можливостей впровадження проєкту є складання SWOT-аналізу (матриці аналізу сильних (Strength) та слабких (Weak) сторін, загроз (Troubles) та можливостей (Opportunities) (таблиця 4.12) на основі

виділених ринкових загроз та можливостей, та сильних і слабких сторін (таблиця 4.11). Перелік ринкових загроз та ринкових можливостей було складено на основі аналізу факторів загроз та факторів можливостей маркетингового середовища. Ринкові загрози та ринкові можливості є наслідками (прогнозованими результатами) впливу факторів, і, на відміну від них, ще не є реалізованими на ринку та мають певну ймовірність здійснення. Наприклад: зниження доходів потенційних споживачів – фактор загрози, на основі якого можна зробити прогноз щодо посилення значущості цінового фактору при виборі товару та відповідно, – цінової конкуренції (а це вже – ринкова загроза).

Таблиця 4.12 - SWOT-аналіз стартап-проекту

Сильні сторони: Ціна Орієнтованість на кінцевого споживача	Слабкі сторони: Кросплатформність програмного продукту
Можливості: Конкуренція Поява нових методів прогнозування Поява нових методів моделювання	Загрози: Невідповідність умовам соціального розвитку Зміна потреб користувачів

На основі SWOT-аналізу було розроблено альтернативи ринкової поведінки (перелік заходів) для виведення стартап-проекту на ринок та орієнтовний оптимальний час їх ринкової реалізації з огляду на потенційні проекти конкурентів, що можуть бути виведені на ринок (див. таблицю 4.9, аналіз потенційних конкурентів). Визначені альтернативи були проаналізовані з точки зору строків та ймовірності отримання ресурсів (таблиця 4.13).

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Безкоштовне розповсюдження створеного програмного продукту	45%	18 місяців
2	Створення програмного продукту з подальшим розповсюдженням за певну оплату	85%	18 місяців
3	Створення веб-сайту, на якому можна буде користуватися програмним продуктом	75%	16 місяців

Після аналізу було обрано альтернативу №2

4.4 Аналіз ринкової стратегії проекту

Розроблення ринкової стратегії першим кроком передбачає визначення стратегії охоплення ринку: було проведено опис цільових груп потенційних споживачів (таблиця 4.14).

За результатами аналізу потенційних груп споживачів було обрано цільові групи, для яких буде запропоновано даний товар, та визначено стратегію охоплення ринку – стратегію диференційованого маркетингу (компанія працює з декількома сегментами).

Таблиця 4.14 - Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готов- ність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Просто та входу у сегмент
1	Компанії (українські та міжнародні) діяльність яких пов'язана з фінансово- економічною сферами.	Висока	Високий	Сильна	Склад- но
2	Приватні підприємства міського та міжнародного рівня, діяльність яких пов'язана з фінансово- економічною сферами.	Висока	Високий	Сильна	Склад- но
3	Приватні підприємства, обласного рівня.	Помірна	Помірний	Помірна	Серед- ня склад- ність

Продовження таблиці 4.14

4	Підприємства регіонального характеру	Помірна	Слабкий	Слабка	Просто
5	ФОП, які діють у фінансово- економічній сферах.	Слабка	Слабкий	Слабка	Просто
Які цільові групи обрано:1,2,3					

Для роботи в обраних сегментах ринку сформовано базову стратегію розвитку (таблиця 4.15).

Таблиця 4.15 - Визначення базової стратегії розвитку

№ п/п	Обрана альтерна тива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспро- можні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
		Визначити потреби кожної з груп, розробити відповідно до них стратегії приваблення клієнтів та комунікації	Цінова політика, універсальність Продукту (миттєве практичне застосування), орієнтованість на кінцевого споживача	Стратегія диференціації

Наступним кроком обрано стратегію конкурентної поведінки (таблиця 4.16)

Таблиця 4.16 - Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохід- цем» на ринку?	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
	«Першопрохо- дець»	Шукати нових	Ні	Стратегія заняття конкурентної ніші

На основі вимог споживачів з обраних сегментів до постачальника (стартап-компанії) та до продукту (див. таблицю 4.5), а також в залежності від обраної базової стратегії розвитку (таблиця 4.15) та стратегії конкурентної поведінки (таблиця 4.16) розроблено стратегію позиціонування (таблиця 4.17), що полягає у формуванні ринкової позиції (комплексу асоціацій), за яким споживачі мають ідентифікувати торгівельну марку/проект.

Результатом виконання підрозділу стала узгоджена система рішень щодо ринкової поведінки стартап-компанії, яка визначає напрями роботи стартап-компанії на ринку.

Таблиця 4.17 - Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартап-проекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
	Легкість розуміння, зручний інтерфейс, надійний, швидкий, точний та достовірний програмний продукт для побудови моделей і прогнозів.	Стратегія диференціації	Позиція на основі порівняння фірми з товарами конкурентів; відмінні особливості споживача	Економія часу; Зручність застосування; Практичність та точність результату

4.5 Розроблення маркетингової програми стартап-проекту

Сформовано маркетингову концепцію товару, який отримає споживач. Для цього у таблиці 4.18 підсумовано результати попереднього аналізу конкурентоспроможності товару. Концепція товару – письмовий опис фізичних та інших характеристик товару, які сприймаються споживачем, і набору вигод, які він обіцяє певній групі споживачів.

Таблиця 4.18 - Визначення ключових переваг концепції потенційного товару

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Швидкість отримання результату	Швидка побудова моделі та створення прогнозу	Відсутність необхідності звертатися до сторонньої особи/компанії для побудови моделі та прогнозу. Дані компанії-користувача, якими оперує програмний продукт, не передаються третім особам, чого вимагає політика безпеки багатьох компаній.
2	Зручність застосу- вання	Не потріб- но мати глибоких знань, для того щоб побудувати модель та спрогнозува ти показники	ПП сам обирає необхідний та оптимальний метод для побудови моделі та прогнозу. Не потрібно мати глибоких знань у прогнозуванні для того, щоб користуватися ПП
3	Практич- ність та точність результату	Користувач отримує точні (з малою похибкою розбіжності) результати.	Користувач на виході роботи ПП отримує модель та прогноз, котрі відповідають необхідним показникам достовірності та точності. Отриманий прогноз можна застосовувати для створення стратегії розвитку підприємства.

Розроблено трирівневу маркетингову модель товару: уточнюється ідея продукту та/або послуги, його фізичні складові, особливості процесу його надання (таблиця 4.19).

1-й рівень. При формуванні задуму товару вирішується питання щодо того, засобом вирішення якої потреби і / або проблеми буде даний товар, яка його основна вигода. Дане питання безпосередньо пов'язане з формуванням технічного завдання в процесі розробки конструкторської документації на виріб.

2-й рівень. Цей рівень являє рішення того, як буде реалізований товар в реальному включає в себе якість, властивості, дизайн, упаковку, ціну.

3-й рівень. Товар з підкріпленням (супроводом) - додаткові послуги та переваги для споживача, що створюються на основі товару за задумом і товару в реальному виконанні (гарантії якості, доставка, умови оплати та ін).

Таблиця 4.19 - Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Зручність та швидкість отримання практичного результату щодо побудови моделі та прогнозування процесів.		
II. Товару реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. функція побудови моделі процесу		
	2. функція побудови прогнозу		
	Якість: достовірність побудови математичної моделі, достовірність побудови прогнозу		
	Пакування: відсутнє		
	Марка: «Stat Forecast»		

Продовження таблиці 4.19

III. Товар із підкріпленням	До продажу: відсутнє
	Після продажу: персональна підтримка в обслуговуванні за додаткову платню.
Вихідний код та математична модель будуть закриті. На ідею зареєстровано патент.	

Після формування маркетингової моделі товару слід відмітити, що проект буде захищено від копіювання. Наступним кроком є визначення цінових меж, якими необхідно керуватись при встановленні ціни на потенційний товар (остаточне визначення ціни відбувається під час фінансово-економічного аналізу проекту), яке передбачає аналіз ціни на товари-аналоги або товари субституту, а також аналіз рівня доходів цільової групи споживачів (таблиця 4.20). Аналіз проведено експертним методом.

Таблиця 4.20 - Визначення меж встановлення ціни

№ п/п	Рівень цін на товари-замінники	Рівень цін на товари-аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
	1800\$	3500\$	У всіх трьох груп високий рівень доходів	Базова покупка 1000\$ Подальша персональна підтримка в обслуговуванні 150\$/місяць

Наступним кроком є визначення оптимальної системи збуту, в межах якого було прийняте рішення (таблиця 4.21).

Таблиця 4.21 - Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
	Цільові клієнти – компанії, які бажають впровадити у своїй роботі сучасні засоби, які допоможуть отримати вигоду та покращити дохідність. Вони цікавляться сучасними розробками та інноваційними рішеннями, тому відвідують конференції, інтернет-конференції, семінари.	Встановлення контактів із споживачами і підтримання їх. Формування попиту і стимулювання збуту. Дослідницька робота зі збору маркетингової інформації. Доробка товару, виходячи з потреб конкретного покупця.	Один (від виробника одразу споживачу)	Прямий канал збуту до споживача, мінімізувати збутові витрати розвитку маркетингового спілкування із споживачем

Останньою складовою маркетингової програми є розроблення концепції маркетингових комунікацій, що спирається на попередньо обрану основу для позиціонування, визначену специфіку поведінки клієнтів (таблиця 4.22).

Таблиця 4.22 – Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
	Цільові клієнти – компанії, які бажають впровадити у своїй роботі сучасні засоби, які допоможуть отримати вигоду та покращити дохідність. Вони цікавляться сучасними розробками та інноваційними рішеннями, тому відвідують конференції, інтернет-конференції.	Конференції, інтернет-конференції, семінари, огляд професійної літератури, інтернет, періодичні видання у різноманітних (профільних) галузях.	Позиція на основі порівняння фірми з товарами конкурентів; відмінності особливості споживача	Створення репутації фірми; збільшення чистого прибутку та рентабельності фірми; збільшення потоків покупців та обсягів продажу; стабілізація обсягів продажу в період зменшення попиту та загального спаду ділової активності.	Шукаєте вірний шлях для розвитку вашої компанії? Досить даремно гаяти час на вгадування вірної стратегії! Користуйтеся «Stat Forecast» і світле майбутнє вашій компанії забезпечено!

Результатом підрозділу стала ринкова (маркетингова) програма, що включає в себе концепції товару, збуту, просування та попередній аналіз можливостей ціноутворення, спирається на цінності та потреби потенційних клієнтів, конкурентні переваги ідеї, стан та динаміку ринкового середовища, в межах якого впроваджено проект, та відповідну обрану альтернативу ринкової поведінки.

Висновки до розділу 4

В даному розділі було проведено аналіз програмного продукту у якості стартап проекту. Можна зазначити, що у проекту є можливість комерціалізації, оскільки ринок потребує якісний продукт, що надає можливість створювати моделі нелінійних нестационарних процесів.

На ринку наявна конкуренція, існує декілька фірм-конкурентів, але їх товар дещо відрізняється, тому вихід на ринок не буде легким і потребує грамотної стратегії виходу. Для впровадження ринкової реалізації проекту слід обрати альтернативу, яка передбачає розробку програмного продукту з подальшим розповсюдженням за певну плату.

Отже, подальший розвиток проекту є доцільним, оскільки він знайде свою цільову аудиторію.

ВИСНОВКИ

У магістерській дисертації розглянуто основні типи нестационарних процесів та їхні характерні особливості. Проведено порівняльний аналіз різних програмних продуктів для статистичної обробки даних, визначено основні переваги та недоліки кожного з них. Здійснено огляд існуючих підходів до моделювання і прогнозування нелінійних нестационарних процесів: параметричних і непараметричних, також розглянуто останні дослідження та розробки в області прогнозування процесів. Запропоновано методику моделювання нелінійних нестационарних процесів, представлених статистичними даними.

У роботі реалізовано моделювання на основі декількох моделей: авторегресійних моделей (АР), авторегресійних моделей з ковзним середнім (АРКС) та методу групового врахування аргументів (МГУА). Наведені методи було практично застосовано для моделювання таких часових рядів, як Google_stocks, Amazon_stocks, CPI_Ukr та CPI_Germany. Встановлено, що мінімального значення середньої квадратичної похибки можна досягти за рахунок зміни порядку регресії та порядку ковзного середнього.

Створено програмний продукт, який дозволяє реалізовувати тест Дікі-Фуллера, вилучати тренди за допомогою різниць часового ряду, будувати графіки функцій автокореляції та часткової автокореляції, створювати ARMA/ARIMA моделі та здійснювати прогнозування.

На основі результатів моделювання і прогнозування за допомогою різних методів можна зробити висновок про те, що не існує єдиного універсального методу. Для отримання найбільш адекватної моделі та якісного прогнозу слід застосовувати різні підходи і враховувати параметри та критерії.

ПЕРЕЛІК ПОСИЛАНЬ

1. Cheng C., Sa-Ngasoongsong A., Beyca O., та ін. Time series forecasting for nonlinear and non-stationary processes: A review and comparative study. *IIE Transactions (Institute of Industrial Engineers)*. 2015. Vol. 47, No. 10. P. 1053–1071.
2. Kaldor N. A Model of the Trade Cycle. *The Economic Journal*. 1940. Vol. 50, No. 197. P. 78–92.
3. Kaldor N. The Irrelevance of Equilibrium Economics. *The Economic Journal*. 1972. Vol. 82, No. 328. P. 1237.
4. Fan J., Yao Q. Nonlinear Time Series: Nonparametric and Parametric Methods: Nonlinear Time Series: Nonparametric and Parametric Methods. New York: Springer, 2003. 569 p.
5. Kumar C., Patel N., Jarudi I. 21st century physics: grand challenges. *Journal of the Federation of American Scientists*. 2003. Vol. 56, No. 2. P. 9–11.
6. Stationary process. URL: https://en.wikipedia.org/wiki/Stationary_process.
7. Бідюк П. І., Романенко В. Д., Тимощук О. Л. Аналіз Часових Рядів: Київ: Політехніка, 2010. 317 с.
8. Heteroskedasticity.
URL: <https://www.investopedia.com/terms/h/heteroskedasticity.asp>.
9. Gooijer J. G. De, Hyndman R. J. 25 Years of Time Series Forecasting. *International Journal of Forecasting*. 2006. Vol. 22, No. 3. P. 443–473.
10. Battaglia F., Protopapas M. K. Time-varying multi-regime models fitting by genetic algorithms. *Journal of Time Series Analysis*. 2011. Vol. 32, No. 3. P. 237–252.
11. Engle R. F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*. 1982. Vol. 50, No. 4. P. 987.
12. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *Journal of*

- Econometrics*. 1986. Vol. 31, No. 3. P. 307–327.
13. Tong H. Non-Linear Time Series: A Dynamical System Approach. Oxford: Clarendon Press, 1990. 564 p.
 14. Yadav P. K., Pope P. F., Paudyal K. Threshold Autoregressive Modeling in Finance: the Price Differences of Equivalent Assets. *Mathematical Finance*. 1994. Vol. 4, No. 2. P. 205–221.
 15. McAleer M., Medeiros M. C. A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. *Journal of Econometrics*. 2008. Vol. 147, No. 1. P. 104–119.
 16. Lundbergh S., Teräsvirta T., Dijk D. Van. Time-varying smooth transition autoregressive models. *Journal of Business and Economic Statistics*. 2003. Vol. 21, No. 1. P. 104–121.
 17. Lineesh M., John C. Analysis of non-stationary time series using wavelet decomposition. *Nature and Science*. 2010. Vol. 8, No. 1. P. 53–59.
 18. Lowe D., Webb A. R. Time series prediction by adaptive networks: a dynamical systems perspective: *IEE Proceedings of Radar and Signal Processing*, 91. P. 17–24.
 19. Kohzadi N., Boyd M., Kermanshahi B. et. al. A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*. 1996. Vol. 10, No. 2. P. 169–181.
 20. Hippert H., Pedreira C., Souza R. C. Neural networks for short-term load forecasting: a review and evaluation: *IEEE Transactions on Power Systems*, 01. P. 44–55.
 21. Park J., Sandberg I. W. Universal approximation using radialbasis- function networks. *Neural Computation*,. 1991. Vol. 3, No. 2. P. 246–257.
 22. Zhang G. P., Rozenberg G., Back T. et. al. Neural networks for time-series forecasting: 2012. 461–477 c.
 23. Lapedes A., Farber R. Nonlinear signal processing using neural networks: prediction and system modeling. San Diego: SciTech Connect, 1987. 52 p.
 24. French M., Krajewski W., Cuykendall R. R. Rainfall forecasting in space and time

- using a neural network. *Journal of Hydrology*. 1992. Vol. 137, No. 1. P. 1–31.
25. Groot C. De, Wuertz D. Analysis of univariate time series with connectionist nets: a case study of two classical examples. *Neurocomputing*. 1991. Vol. 3, No. 4. P. 177–192.
 26. Grudnitski G., Osburn L. Forecasting S&P and gold futures prices: an application of neural networks. *Journal of Futures Markets*. 1993. Vol. 13, No. 6. P. 631–643.
 27. Kuan C. M., Liu T. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*. 1995. Vol. 10, No. 4. P. 347–364.
 28. Menezes J. M. P., Barreto G. A. Long-term time series prediction with the NARX network: an empirical evaluation. *Neurocomputing*. 2008. Vol. 71, No. 16–18. P. 3335–3343.
 29. Yee P., Haykin. Adaptive regularized radial basis function network for nonlinear, nonstationary time series prediction: *IEEE Transactions on Signal Processing*, 99. P. 2503–2521.
 30. Chen S., Cowan C. F. N., Grant P. M. Orthogonal least squares learning algorithm for radial basis function networks: *IEEE Transactions on Neural Networks*, 91. P. 302–309.
 31. Chen S., Billings S. A., Grant P. M. Recursive hybrid algorithm for non-linear system identification using radial basis function networks. *International Journal of Control*. 1992. Vol. 55, No. 5. P. 1051–1070.
 32. Barreto G., Hammer B., Hitzler P. Time Series prediction with the self-organizing map: a review perspectives of neural-symbolic integration. *Perspectives of Neural-Symbolic Integration*. 2007. P. 135–158.
 33. Nayak P. C., Sudheer K. P., Rangan D. M. et. al.. A neuro-fuzzy computing technique for modeling hydrological time series. *Journal of Hydrology*. 2004. Vol. 291, No. 1–2. P. 52–66.
 34. Smola A., Scholkopf B. A tutorial on support vector regression. *Statistics and Computing*. 2004. Vol. 14, No. 3. P. 199–222.
 35. Balabin R. M., Lomakina E. I. Support vector machine regression (SVR/LS-

- SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data. *Analyst*. 2011. Vol. 136, No. 8. P. 1703–1712.
36. Mukherjee S., Osuna E., Girosi F. Nonlinear prediction of chaotic time series using support vector machines: *Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing*, 97. P. 511–520.
 37. Lau K. W., Wu Q. H. Local prediction of non-linear time series using support vector regression. *Pattern Recognition*. 2008. Vol. 41, No. 5. P. 1539–1547.
 38. Khemchandani R., Chandra S. Regularized least squares fuzzy support vector regression for financial time series forecasting. *Expert Systems with Applications*. 2009. Vol. 36, No. 1. P. 132–138.
 39. Rabiner L. Tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*. 1989. Vol. 77, No. 2. P. 257–286.
 40. Gonzalez A. M., Roque A. M. S., Garcia-Gonzalez J. Modeling and forecasting electricity prices with input/output hidden Markov models. *IEEE Transactions on Power Systems*, 2005. Vol. 20, No.1. P. 13–24.
 41. Yankov D., DeCoste D., Keogh E. Ensembles of nearest neighbor forecasts. *Lecture Notes in Computer Science*. 2006. Vol. 4212. P. 545–556.
 42. Hamid N. Z. A., Noorani M. S. M. Modeling of prediction system: an application of nearest neighbor approach to chaotic data. *Applied Mathematics and Computational Intelligence*. 2013. Vol. 2. P. 137–148.
 43. Gershman S. J., Blei D. M. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*. 2012. Vol. 56, No. 1. P. 1–12.
 44. Rasmussen C. E., Williams C. K. I. Gaussian Process for Machine Learning. Cambridge: MIT Press, 2006. 266 p.
 45. Girard A., Rasmussen C. E. et. al. Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting: Proceedings of the 15th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2002. P. 545–552.
 46. Frei M. G., Osorio I. Intrinsic time-scale decomposition: time–frequency–energy

- analysis and real-time filtering of nonstationary signals: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*. 2007. Vol. 463, No. 2078. P. 321–342.
47. Zhang G. P., Qi M. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*. 2005. Vol. 160, No. 2. P. 501–514.
48. Agung I. G. N. Time Series Data Analysis Using EViews. Singapore: Wiley, 2009. 635 p.
49. EViews Fundamentals.
URL: http://www.eviews.com/help/helpintro.html#page/content%2Fsec_fund.html%2
50. SAS/STAT Software Regression Analysis.
URL: <https://support.sas.com/rnd/app/stat/procedures/Regression.html>.
51. Scikit-learn Machine Learning in Python. URL: <https://scikit-learn.org/stable/>.
52. Statsmodels Python. URL: <https://www.statsmodels.org/stable/index.html>.
53. Носко В. П. Эконометрика Введение в регрессионный анализ временных рядов. Москва: МФТИ, 2002. 273 с.
54. Ivakhnenko A., Ivakhnenko G. The review of problems solvable by algorithms of the group method of data handling (GMDH). *Pattern Recognition And Image Analysis*. 1995. Vol. 5. P. 527–535.
55. Степашко В. С., Ивахненко А. Г. Помехоустойчивость моделирования. Киев: Наукова думка, 1985. 216 с.
56. Бідюк П. І., Терентьев О. М., Просянкін-Жарова Т. І. Прикладна статистика: Вінниця: ПП “ТД”Едельвейс і К”, 2013. 304 с.
57. White H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica*. 1980. Vol. 48, No. 4. P. 817–838.

ДОДАТОК А ЛІСТИНГ ПРОГРАМИ

```

import pandas as kunfu
import numpy as dragon
import pylab as p
import matplotlib.pyplot as plot
from collections import Counter
import re

#importing packages for the prediction of time-series data
import statsmodels.api as sm
import statsmodels.tsa.api as smt
import statsmodels.formula.api as smf

from sklearn.metrics import mean_squared_error
data = train['Close']
Date1 = train['Date']
print(train.info())
train1 = train[['Date','Close']]
train1.head()
# Setting the Date as Index
train2 = train1.set_index('Date')
train2.sort_index(inplace=True)
print (type(train2))
print (train2.head())
plot.plot(train2)
plot.xlabel('Date', fontsize=12)
plot.ylabel('Price in USD', fontsize=12)
plot.title("Closing price distribution of data series", fontsize=15)
from statsmodels.tsa.stattools import adfuller

def test_stationarity(x):

    #Determining rolling statistics

```

```
rolmean = x.rolling(window=22,center=False).mean()
```

```
rolstd = x.rolling(window=12,center=False).std()
```

```
#Plot rolling statistics:
```

```
orig = plot.plot(x, color='blue',label='Original')
```

```
mean = plot.plot(rolmean, color='red', label='Rolling Mean')
```

```
std = plot.plot(rolstd, color='black', label = 'Rolling Std')
```

```
plot.legend(loc='best')
```

```
plot.title('Rolling Mean & Standard Deviation')
```

```
plot.show(block=False)
```

```
#Perform Dickey Fuller test
```

```
result=adfuller(x)
```

```
print('ADF Stastistic: %f'%result[0])
```

```
print('p-value: %f'%result[1])
```

```
pvalue=result[1]
```

```
for key,value in result[4].items():
```

```
    if result[0]>value:
```

```
        print("The graph is non stationery")
```

```
        break
```

```
    else:
```

```
        print("The graph is stationery")
```

```
        break;
```

```
print('Critical values:')
```

```
for key,value in result[4].items():
```

```
    print("\t%s: %.3f ' % (key, value))
```

```
ts = train2['Close']
```

```
test_stationarity(ts)
```

```
ts_log = dragon.log(ts)
```

```
plot.plot(ts_log,color="green")
```

```
plot.show()
```

```
test_stationarity(ts_log)
```

```

ts_log = dragon.log(ts)
plot.plot(ts_log,color="green")
plot.show()

test_stationarity(ts_log)
# Naive decomposition of our Time Series as explained above
from statsmodels.tsa.seasonal import seasonal_decompose
decomposition = seasonal_decompose(ts_log, model='multiplicative',freq = 7)

trend = decomposition.trend
seasonal = decomposition.seasonal
residual = decomposition.resid

plot.subplot(411)
plot.title('Observed = Trend + Seasonality + Residuals')
plot.plot(ts_log,label='Observed')
plot.legend(loc='best')
plot.subplot(412)
plot.plot(trend, label='Trend')
plot.legend(loc='best')
plot.subplot(413)
plot.plot(seasonal,label='Seasonality')
plot.legend(loc='best')
plot.subplot(414)
plot.plot(residual, label='Residuals')
plot.legend(loc='best')
plot.tight_layout()
plot.show()
ts_log_decompose = residual
#ts_log_decompose.fillna(inplace=True)
#test_stationarity(ts_log_decompose)
ts_log_diff = ts_log - ts_log.shift()
plot.plot(ts_log_diff)
plot.show()
ts_log_diff.dropna(inplace=True)

```

```

test_stationarity(ts_log_diff)

#ACF and PACF plots
from statsmodels.tsa.stattools import acf, pacf
from statsmodels.tsa.arima_model import ARIMA

lag_acf = acf(ts_log_diff, nlags=10)
lag_pacf = pacf(ts_log_diff, nlags=10, method='ols')

#Plot ACF:
plot.subplot(1,1,1)

plot.plot(lag_acf)
#print(lag_acf[0.5])

plot.axhline(y=0,linestyle='--',color='g')
plot.title('Autocorrelation Function')
plot.show()

#Plot PACF:
plot.subplot(1,1,1)
plot.plot(lag_pacf)

plot.axhline(y=0,linestyle='--',color='green')
plot.title('Partial Autocorrelation Function ')
plot.tight_layout()
plot.show()

# follow lag
model = ARIMA(ts_log, order=(4,1,0))
results_ARIMA = model.fit(dis=-1)
plot.plot(ts_log_diff)
plot.plot(results_ARIMA.fittedvalues, color='red')
plot.title('RSS: %.7f'% sum((results_ARIMA.fittedvalues-ts_log_diff)**2))
plot.show()

# follow error
model = ARIMA(ts_log, order=(0,1,0))

```

```

results_MA = model.fit(dis=-1)
plot.plot(ts_log_diff)
plot.plot(results_MA.fittedvalues, color='red')
plot.title('RSS: %.7f% sum((results_MA.fittedvalues-ts_log_diff)**2))
plot.show()
print(results_MA.summary())
from statsmodels.tsa.arima_model import ARIMA
model = ARIMA(ts_log, order=(4,1,0))
results_ARIMA = model.fit(dis=-1)
plot.plot(ts_log_diff)
plot.plot(results_ARIMA.fittedvalues, color='red')
plot.title('RSS: %.7f% sum((results_ARIMA.fittedvalues-ts_log_diff)**2))
plot.show()
from statsmodels.tsa.arima_model import ARIMA
from pandas import DataFrame
print(results_ARIMA.summary())
size = int(len(ts_log)-100)
train_arima, test_arima = ts_log[0:size], ts_log[size:len(ts_log)]
history = [x for x in train_arima]
predictions = list()
originals = list()
error_list = list()

print('Printing Predicted vs Expected Values...')
print('\n')
for t in range(len(test_arima)):
    model = ARIMA(history, order=(2, 1, 0))
    model_fit = model.fit(dis=-1)

    output = model_fit.forecast()

    pred_value = output[0]

    original_value = test_arima[t]
    history.append(original_value)

```

```

pred_value = dragon.exp(pred_value)

original_value = dragon.exp(original_value)

error = ((abs(pred_value - original_value)) / original_value) * 100
error_list.append(error)
print('predicted = %f, expected = %f, error = %f ' % (pred_value, original_value, error), '%')

predictions.append(float(pred_value))
originals.append(float(original_value))

#error = mean_squared_error(dragon.exp(yhat), dragon.exp(obs))
#print('mean_squared_error : ', error)
print("\n Means Error in Predicting Test Case Articles : %f ' %
(sum(error_list)/float(len(error_list))), '%')
plot.figure(figsize=(8, 6))
test_day = [t for t in range(len(test_arima))]
labels={'Original','Predicted'}
plot.plot(test_day, predictions, color= 'green')
plot.plot(test_day, originals, color = 'orange')
plot.title('Expected Vs Predicted Views Forecasting')
plot.xlabel('Day')
plot.ylabel('Closing Price')
plot.legend(labels)
plot.show()
model3=sm.OLS(originals,predictions)
result=model3.fit()
print(result.summary())

```